University of Alberta

A Decision Support System for Variogram Calculation

by

Seyed Hadi Derakhshan

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Mining Engineering

Department of Civil and Environmental Engineering

Edmonton, Alberta
Fall 2007

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-33232-0
Our file Notre référence
ISBN: 978-0-494-33232-0

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Dedicated to

my father, Mohammad, and my mother, Ashraf,

with all my love

Abstract

The variogram model used in estimation and/or simulation is unquestionably important. Quantifying and assessing the variogram model with less uncertainty result in reasonable models and simulations. In real case when the true variogram is not known, the only thing that can help us to quantify the true variogram with associated uncertainty is our data. In the case of regularly gridded data calculating experimental variogram is much easier than the case of irregular data set; in this latter case some tolerance parameters should be defined to have enough data pairs for calculating a reliable variogram.

These tolerance parameters can be optimized by minimizing a well defined penalty function that accounts for the difference in the fitted variogram and the assumed to be true variogram. This work gives a reasonable tool to assess the experimental variogram with less uncertainty.

Acknowledgements

I would like to express my deepest appreciation to my research advisors Dr. Oy Leuangthong and Dr. Clayton V. Deutsch for their support, encouragement and guidance throughout this work.

Expressed gratitude is given to the sponsors of the Centre for Computational Geostatistics (CCG) and the University of Alberta for their support.

I would like to extend my gratitude to my best friend Mehran for his friendship and support during my undergraduate and graduate studies.

Last but not the least; my unending appreciation goes to my parents, Mohammad and Ashraf, my brothers Mehdi and Amir and my lovely sister Shirin. Their love, support, care, advice, encouragement and patience can not be quantified.

Table of Contents

CHAPTER 1: INTRODUCTION.....1

1.1 Background on Variogram.....1

1.2 Introductory example.....5

1.3 Outline of the Thesis.....8

CHAPTER 2: VARIOGRAM CALCULATIONS.....9

2.1 Spatial relationship.....9

2.2 Calculating and fitting the experimental variogram.....16

CHAPTER 3: IMPROVING VARIOGRAM CALCULATIONS.....20

3.1 Methodology and principles.....20

3.2 Implementation aspects.....22

3.2.1 Optimizing tolerance parameters.....22

3.2.1.1 Tolerance ratio in 2D.....26

3.2.1.2 Tolerance ratio in 3D.....28

3.2.1.3 Methodology.....31

3.2.2 Calculating the tolerance correction factor.....36

3.2.2.1 Tolerance correction factor in 2D without bandwidth.....36

3.2.2.2 Tolerance correction factor in 2D with bandwidth.....41

3.2.2.3 Tolerance correction factor in 3D without bandwidth.....43

3.2.2.4 Tolerance correction factor in 3D with bandwidth.....45

3.2.3 Rescaling the experimental variogram points.....47

CHAPTER 4: APPLICATION.....49

4.1 Synthetic case.....49

4.2 Real example.....64

CHAPTER 5: CONCLUSION AND FUTURE WORK.....72

5.1 Optimal tolerance parameters for variogram calculation 72
5.2 Future work.....73

BIBLIOGRAPHY

APPENDIX A: Theoretical derivations for some of the formulas presented.....A-1

List of Figures

Figure 1.1. Moment of inertia interpretation of the variogram based on an h-scatterplot; (Redrawn from Goovaerts, 1997).....2

Figure 1.2. Different Lag Distances in the Case of Regular Spaced Data: **h**=1 taken vertically yields 40 pairs (top left); **h**=1 taken horizontally results in 42 pairs (top right); **h**=2 in the horizontal direction will give 36 pairs (bottom left); and **h**=3 in the horizontal direction results in 30 pairs (bottom right).....4

Figure 1.3. Locations of synthetic data set.....6

Figure 1.4. Probability density function and cumulative distribution function for the minimum distance from wells.....7

Figure 1.5. Experimental variogram points for different unit lag distances and lag tolerances, the tolerance ratio is the ratio between the lag tolerance and the unit lag distance.....7

Figure 3.1. Fitted variogram (the one which has large nugget effect) versus the reference variogram (the one which has small nugget effect) and their difference (fitted minus reference).....22

Figure 3.2. Lag separation distance, h, lag tolerance, h_{tol}, azimuth tolerance, a_{tol} and bandwidth, b.....24

Figure 3.3. Lag separation distance, h, lag tolerance, h_{tol}, angle tolerance, a_{tol} and bandwidth, b in three dimensional case-vertical lag distance.....24

Figure 3.4. Lag separation distance, h; lag tolerance, h_{tol}; vertical angle tolerance, a^{v}_{tol}; horizontal angle tolerance, a^{h}_{tol}; vertical bandwidth, b_{ver} and horizontal bandwidth, b_{hor} in three dimensional case-horizontal lag distance.....25

Figure 3.5. Tolerance ratio for two dimensional case.....26

Figure 3.6. Anisotropic ellipse in 2D case when there is no bandwidth.....37

Figure 3.7. Anisotropic ellipse in 2D case with bandwidth for maximum direction of continuity.....42

Figure 4.1. Map of reference model generated by `sgsim`.....49

Figure 4.2. Variogram reproduction for the reference model generated by `sgsim`.....50

Figure 4.3. Locations of 200 randomly picked points from reference model.....51

Figure 4.4. Reference distribution (left) and the distribution of the 200 randomly picked points (right).....51

Figure 4.5. The Reference variogram (solid line) and the experimental variogram points based on 200 picked points (bullets).....52

Figure 4.6. Different realizations for variogram at each lag accompanying by the probability density function at **h**=40.....53

Figure 4.7. Probability density function of the variogram values at each lag distances (see Figure 4.6), for each histogram a box plot is shown, the black dot is the reference variogram value.....54

Figure 4.8. Variance of the variogram versus the lag distance.....55

Figure 4.9. Calculated experimental variogram for 100 realizations from Spatial Bootstrap at h=8 and h_{tol}=4.48 (left) and the corresponding fitted variogram values (right).....56

Figure 4.10. Variance of the experimental variogram (dash line) and the variance of the fitted variogram (solid line) versus the lag distance.....56

Figure 4.11. Final penalty as a function of both lag tolerance and lag separation distance.....57

Figure 4.12. Conditional expectation of the average penalty versus the unit lag distance (left) and the tolerance ratio (right).....57

Figure 4.13. Map of reference model generated by `sgsim` for field size of 512.....58

Figure 4.14. Variogram reproduction for the reference model with field size of 512.....59

Figure 4.15. Locations of different number of randomly picked points from reference model with field size of 512.....59

Figure 4.16. Locations of 256 of randomly picked points for different realizations (256 data locations are different) from the same reference model with field size of 1024.....60

Figure 4.17. Penalty function in the case of different number of data and a fixed field size.....61

Figure 4.18. Conditional expectation of the average penalty versus the unit lag distance (left) and the tolerance ratio (right) for different number of picked points and a fixed field size of 1024, the solid line is the average over different realizations (the dashed lines).....62

Figure 4.19. Fitted variogram (dashed line) versus the reference variogram (solid line) in the case of different number of data (n=128, 256, 512 and 1024) and a fixed field size ($|A| = 512$) using optimal tolerance parameters, the solid line is the reference variogram, the dashed line is the fitted variogram and the dots are the experimental variogram points.....63

Figure 4.20. Optimal tolerance ratio versus the number of data.....63

Figure 4.21. Location map of the 2D porosity data.....64

Figure 4.22. Location map of the 2D porosity data.....64

Figure 4.23. Distribution of the minimum distance between wells.....65

Figure 4.24. Plot of penalty function as a function of unit lag distance and the tolerance ratio.....66

Figure 4.25. Plot of penalty function as a function of angle tolerance for Amoco 2D data set.....68

Figure 4.26. Plot of penalty function as a function of unit lag distance and the tolerance ratio for different nugget effect values for reference variogram model used in penalty function calculation for Amoco 2D data; the reference variogram range is set to be 30 percent of the field size.....69

Figure 4.27. Plot of penalty function as a function of unit lag distance and the tolerance ratio for different reference variogram range for variogram model used in penalty function calculation for Amoco 2D data; the nugget effect is set to be 0 percent.....70

Figure 4.28. Plot of penalty function as a function of vertical bandwidth for Amoco 3D data set, the vertical bandwidth is a value between 0 and 1 because it is in proportional stratigraphic unit which is a fraction.....71

List of Nomenclature

θ	A set of tolerance parameters
a_{tol}	Angle tolerance
$a_{tol,\max}^{app}$	Apparent angle tolerance for maximum direction in 2D case
$a_{tol,\min}^{app}$	Apparent angle tolerance for minimum direction in 2D
$a_{tol,\max}^{v,app}$	Apparent vertical angle tolerance for maximum direction
$a_{tol,\min}^{v,app}$	Apparent vertical angle tolerance for minimum direction
$a_{tol,\max}^{h,app}$	Apparent horizontal angle tolerance for maximum direction
$a_{tol,\min}^{h,app}$	Apparent horizontal angle tolerance for minimum direction
$\overline{p}(\theta)$	Average of the calculated penalties over different realizations
b	Bandwidth, b
$\hat{\gamma}(\mathbf{h})$	Calculated experimental semivariogram
$C(\mathbf{h})$	Covariance at lag separation vector of $\mathbf{h}$
A	Domain of study
ω	Eccentricity of the anisotropic ellipse
$a_{h-\max}^{*}$	Estimated maximum range of correlation in horizontal plane
$a_{h-\min}^{*}$	Estimated minimum range of correlation in horizontal plane
a_{ver}^{*}	Estimated Vertical range of correlation

$E\{X\}$	Expected value of random variable X
$\gamma^{fit}_{i_{real}}(\mathbf{h};\theta)$	Fitted variogram of the experimental variogram points based on the t tolerance parameters θ.
$\gamma_{GR}(\mathbf{h})$	General relative variogram
a^{h}_{tol}	Horizontal angle tolerance
b_{hor}	Horizontal bandwidth
$F(\alpha,\omega)$	Incomplete Legendre elliptic integral of the 1st kind
$i(\mathbf{u},z_k)$	Indicator transform values (0 or 1)
$wt(\mathbf{h})$	Inverse distance weight associated to each lag distance
$f_{\max}$	Lag distance correction factor in maximum direction
$f_{\min}$	Lag distance correction factor in minimum direction
f_{ver}	Lag distance correction factor in vertical direction
k	Lag number
h	Lag separation distance (scalar)
$\mathbf{h}$	Lag separation vector (vector)
h_{tol}	Lag tolerance (scalar)
$m_{+\mathbf{h}}$	Local mean at head value
$m_{-\mathbf{h}}$	Local mean at tail value
m	Mean of the data
d_k	Mean of the separation distance between the data points of the k^{th} class
$d_{i_{data}}$	Minimum distance between the data points which corresponds to each data location

n_h	Number of different unit lag distances
n_{tol}	Number of different lag tolerances
$C_{NE}(\mathbf{h})$	Non-ergodic covariance
$\gamma_{NE}(\mathbf{h})$	Non-ergodic variogram
$\rho_{NE}(\mathbf{h})$	Non-ergodic correlogram
x_{NS}	Normal score value
$N(\mathbf{h})$	Number of pairs of data separated by a vector **h**
n_{realz}	Number of realizations
$\gamma_{PR}(\mathbf{h})$	Pairwise relative variogram
$p_{i_{realz}}(\theta)$	Penalty function for the i_{realz} realization and set of tolerance parameters
p	Power transform exponent, $(0 < p \leq 1)$
$Z(\mathbf{u})$	Random variable at location **u**
RV	Random variable
$\{Z(\mathbf{u}); \mathbf{u} \in A\}$	Random function
RF	Random function
R_i	Rank of the i^{th} value in sequence (ascending order)
i_{realz}	Realization number
a	Reference variogram range
$\gamma^{ref}(\mathbf{h})$	Reference variogram
$\gamma(\mathbf{h})$	Semivariogram at lag separation vector of **h**
θ	Set of tolerance parameters for variogram calculation

ε	Small number
$\sigma_{+\mathbf{h}}$	Standard deviation at head value
$\sigma_{-\mathbf{h}}$	Standard deviation at tail value
σ	Standard deviation of the data
σ^2	Variance of the data
$C(0)$	Variance of the data
w	Variogram exponent ($0 < w \leq 2$)
$\gamma_L(\mathbf{h})$	Variogram of the log-transformed data values
$\gamma_{PT}(\mathbf{h})$	Variogram of the power transformed data values
$\mathbf{u}$	Vector of location; $\mathbf{u} \in A$
a^v_{tol}	Vertical angle tolerance
b_{ver}	Vertical bandwidth
a_{ver}	Vertical range of correlation
$a_{h-\max}$	True maximum range of correlation in horizontal plane
$a_{h-\min}$	True minimum range of correlation in horizontal plane

Chapter 1

Introduction

One of the simple applications of geostatistics in earth sciences data is to estimate the value of the spatial variable at unsampled location. This estimation at unsampled location acts as the basis for other geostatistical techniques (such as uncertainty quantification and simulation). It uses spatial correlation of the variable over the whole domain of study.

Assume that $\{Z(\mathbf{u}); \mathbf{u} \in A\}$ is a random function model (where $Z(\mathbf{u})$ is a random variable) which represents the value of the variable at all of the locations of interest, $\mathbf{u}$, within the domain A. In reality we do not know for certain $\{Z(\mathbf{u}); \mathbf{u} \in A\}$. Our goal is to characterize it with the associated uncertainty. What is available to us in reality is a data set which represents the variable at selected locations in the domain.

Geostatistics helps to characterize the full model from the available data with associated uncertainty. The variogram is a spatial correlation function that is later used to estimate the random function at unsampled location. It gives information about how dissimilar the value of the variable at one location is to another location separated by a lag vector $\mathbf{h}$. There always exists a true variogram but with real data, it is unknown. In real case we have to use the data to estimate the variogram.

1.1 Background on Variogram

The variogram is a two-point statistic that spatially relates two random variables (RV), $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$:

$$2\gamma(\mathbf{h}) = E\left\{[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]^2\right\} \quad \text{.....(1.1)}$$

where **u** and **h** are location and lag vectors, respectively, in domain A. Matheron (1965) first proposed a method-of-moments approach to approximate the variogram:

$$2\hat{\gamma}(\mathbf{h}_k) = \frac{1}{N(\mathbf{h}_k)} \sum_{i=1}^{N(\mathbf{h}_k)} \left[z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h}_k)\right]^2 \quad \text{.....(1.2)}$$

where $N(\mathbf{h}_k)$ is the number of pairs of data separated by a vector $\mathbf{h}_k$, k is the lag number which is defined based on some tolerance parameters that will be explained later, k=1,…,K lags. $\mathbf{h}_k$ is the mean of the separation distance between the data points of the k^{th} lag. The experimental variogram can be calculated for all of the directions but practically it is calculated for principal directions of continuity (major, minor and vertical).

One can relate each point on a variogram plot to an **h**-scatterplot, which shows all possible pairs of data values whose locations are separated by a certain distance vector **h**. Journel (1989) described the calculation of the variogram from this **h**-scatterplot as calculating the moment of inertia about the 45° line (see Figure 1.1).

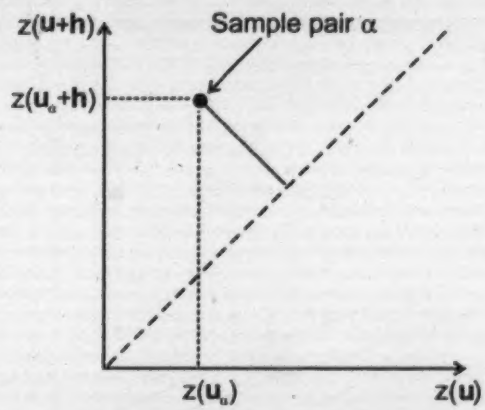


Figure 1.1. Moment of inertia interpretation of the variogram based on an **h**-scatterplot. (Redrawn from Goovaerts, 1997)

Based on the distribution of the cloud of points on an **h**-scatterplot, we can tell how similar the data values are over a certain distance in a particular direction (as defined by the lag vector **h**). If the data values at locations separated by **h** are similar, the pairs will plot close to a 45° line. We naturally expect that this cloud of points will show little

dispersion at small lag distances, but as lag distance increases, this cloud of paired values is expected to increase in dispersion. This notion of dissimilarity (or dispersion) is neatly captured by the variogram.

The number of pairs available for computing the variogram depends on the lag distance. For regularly spaced samples, as the lag separation gets larger there are fewer points, so the method-of-moments approximation for the variogram is less precise at larger lag distances. If there are n observed data, then there are $n(n-1)/2$ unique pairs of observations taken over all possible lag distances. Thus, even a data set of moderate size generates a large number of pairs. Figure 1.2 shows a few different lag distances in the case of regular spaced data for calculating the experimental variogram. We can see that depending on the direction, the lag spacing considered, and the size of the regular grid, the number of pairs used for calculating the variogram can be quite different.

In practice, data are rarely exactly regularly spaced. Sampling campaigns may target nominal drillhole/well spacing; however, certain regions of the deposit/reservoir are inevitably more densely drilled as they provide more information about the available resource. As such, real data are irregularly spaced and the paired information used in calculating the experimental variogram is based on approximate lag separation distances.

Therefore in calculating the experimental variogram in the real case where the data are not on a regular grid some tolerance parameters should be defined to have enough data for calculating reliable variogram with more data pairs (Deutsch and Journel, 1998).

Even after the variogram is numerically calculated, we must still fit the experimental points with a positive semi-definite variogram model. This model is then used in subsequent estimation and/or simulation. Theoretically, we are not constrained to consider any set of models so long as positive semi-definiteness of the resulting model is ensured. Practically, this can be quite prohibitive given the challenges associated to validating that this positive semi-definiteness condition is guaranteed for all directions and all distances. As a result, there are a set of theoretically validated models that are widely adopted including the nugget, spherical, exponential and Gaussian models. These can be linearly combined in an infinite number of ways to fit most experimental

variograms. Gringarten and Deutsch (2001) provide an extensive discussion on variogram interpretation and some guidelines on variogram modelling.

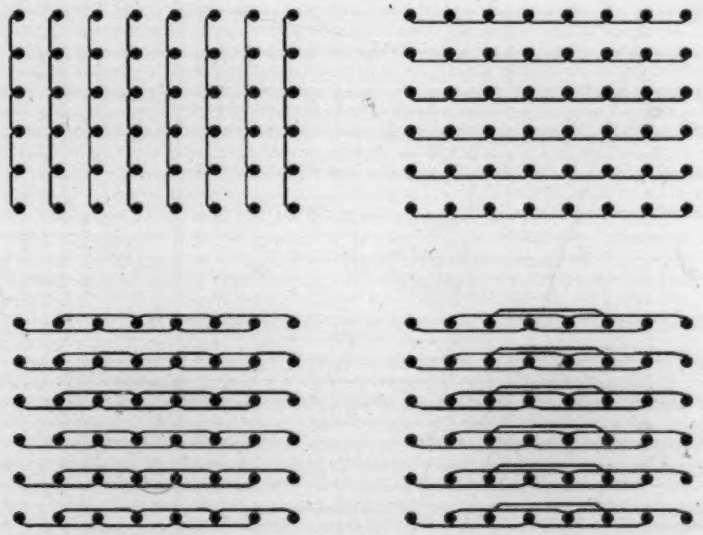


Figure 1.2. Different lag distances in the case of regular spaced data: **h**=1 taken vertically yields 40 pairs (top left); **h**=1 taken horizontally results in 42 pairs (top right); **h**=2 in the horizontal direction will give 36 pairs (bottom left); and **h**=3 in the horizontal direction results in 30 pairs (bottom right).

Of course, the uncertainty in calculating an experimental variogram is carried forward and somehow resolved by the user when the experimental points are fit with a variogram model. The variogram model is required for all distance and direction vectors **h**. The experimental variogram calculation gives variogram values at specific lag distances and directions (usually along the principal directions of continuity), (Deutsch 2002). To build the geostatistical models, we need to have a variogram function for all lag distances and directions. There are some variogram models commonly used in practice, i.e. the nugget effect, spherical, exponential, Gaussian, etc (Deutsch 2002). These models can be used to fit the experimental variogram values.

Specifically there are some key components of the fit that are important (Goovaerts, 1997):

- Although the value of the variogram for **h**=0 is strictly zero, short scale variability may cause sample values separated by extremely small distances (lag) to be quite dissimilar. This results in an apparent vertical intercept on the variogram plot that is often referred to as the nugget effect.
- For a stationary random function, the limit of dissimilarity or the variogram value at which the variogram points appear to converge to at large lag distances is referred to as the sill. We can also interpret the sill as the value at which paired data are no longer correlated to each other, or $C(\mathbf{h})=0$ where $C(\mathbf{h})$ is the covariance of pairs of data separated by **h**. The well established relationship between the variogram, covariance and variance, $\gamma(\mathbf{h})=C(\mathbf{0})-C(\mathbf{h})$, where $C(\mathbf{0})$ represents the variance, demonstrates that the sill of the variogram is equivalent to the variance of the data:

$$C(\mathbf{0})=\sigma^2$$

- The range is the lag distance at or near which the variogram reaches the sill; beyond that distance the corresponding correlation coefficient is zero.

It should be noted that throughout this thesis first and second order stationarity is assumed. These two types of stationarity assumption are defined in chapter 2. There is also a brief explanation of decision of stationarity in chapter 2.

1.2 Introductory example

In this section an introductory example will be presented to show the importance of defining the right parameters for variogram calculation. To understand better the problem for optimizing and defining the reasonable tolerance parameters a synthetic data set is considered with a known variogram model. To have this synthetic data set an unconditional sequential Gaussian simulation is performed with a known isotropic

variogram with a single spherical structure with nugget effect of 0.05 and with field size of 1024x1024, then 128 data points are picked randomly from this simulated field. The locations of the data points are shown in Figure 1.3.

To get a range for reasonable values for lag separation distance the distribution of the distance to nearest sample can be considered. Figure 1.4 shows the probability density function and cumulative distribution function of the minimum distance between data points. It shows that between the unit lag distance of 30 and 40, more data pairs can be found, but this tolerance parameter should be optimized.

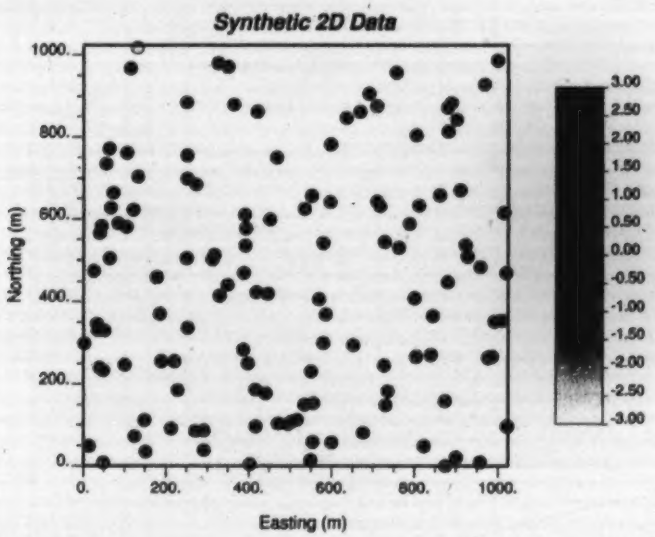


Figure 1.3. Locations of synthetic data set, the values are in Gaussian units

Figure 1.5 shows different experimental variogram plots for different tolerance parameters. Each of these cases shows different variogram from a unique data set. Choosing optimal tolerance parameters will affect on variogram modelling and choosing a right variogram model.

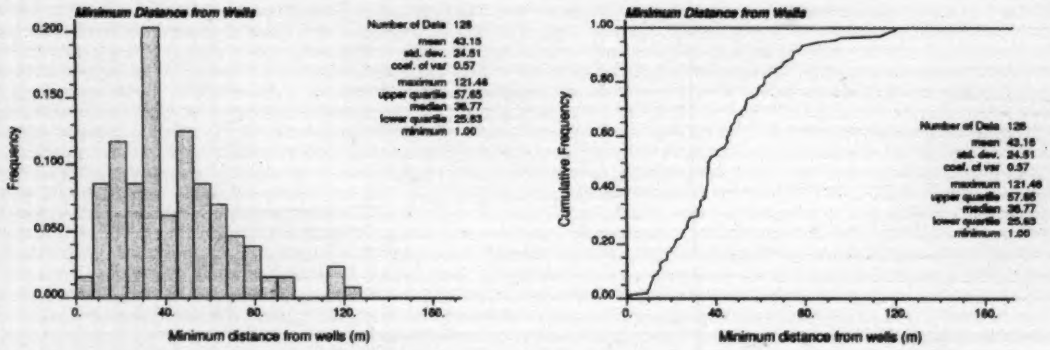


Figure 1.4. Probability density function and cumulative distribution function for the minimum distance from wells

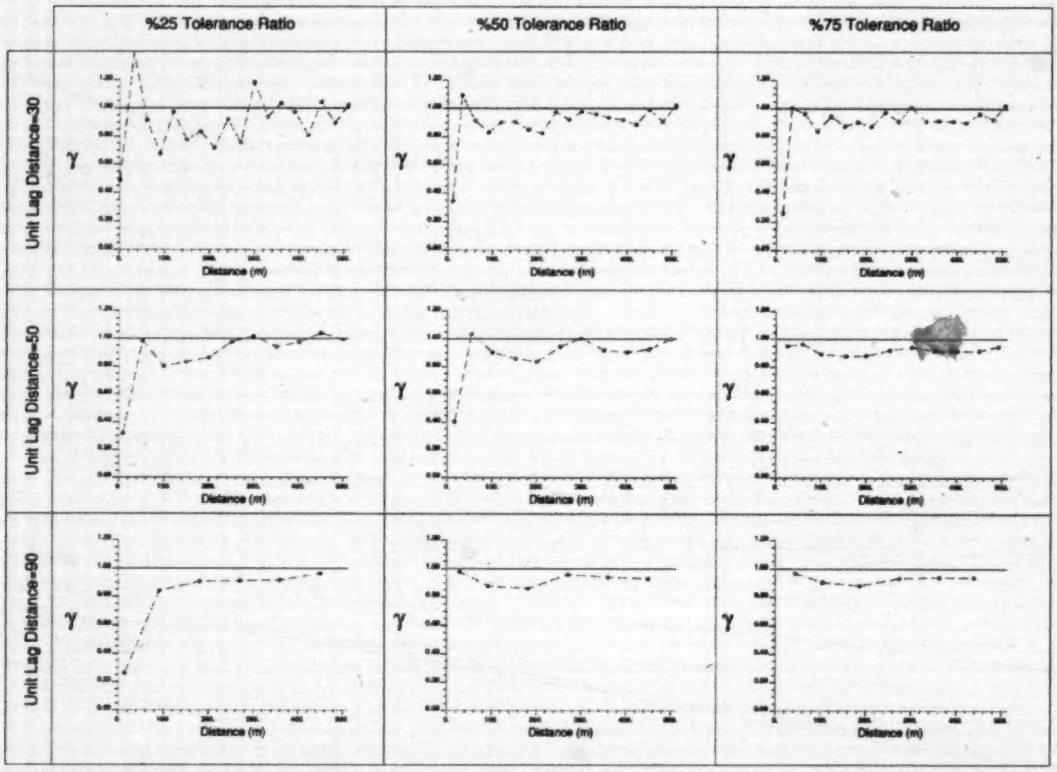


Figure 1.5. Experimental variogram points for different unit lag distances and lag tolerances, the tolerance ratio is the ratio between the lag tolerance and the unit lag distance

1.3 Outline of the thesis

In this study a decision support system will be introduced to improve variogram calculation and modelling. The goal is to model the experimental variogram in such a way that it gets closer to the true variogram which is in almost all of the cases is unknown.

In chapter 2 the theoretical issues related to the variogram calculations and fitting will be discussed. The focus is more on defining all of the parameters that affect the calculation of the variogram. These parameters are directions and tolerance parameters which consist of unit lag distance, lag tolerance, angle tolerance (vertical, horizontal) and the bandwidth (vertical, horizontal). The rest of the chapter is an introduction to variogram fitting, it is considered in this chapter because the following chapters largely use automatic variogram fitting.

In chapter 3 the emphasis is on defining the optimal tolerance parameters based on the data values and their locations. The idea of optimizing the tolerance parameters will be presented in detail, the proposed methodology is to define a penalty function which is function of these parameters and try to find the point where the penalty is minimized with respect to the variogram calculation parameters. The implementation aspects regarding two and three dimensional cases will be discussed.

In chapter 4 a case study will be presented. Real 2D and 3D data sets are used to test the optimization technique and quantify the optimal tolerance parameters.

In chapter 5, the conclusions and future work for the proposed methodology will be discussed.

Chapter 2

Variogram Calculations

Throughout this thesis, **h** is the lag distance which is a vector, h is the unit lag distance and h_{tol} is the unit lag tolerance. Both of the unit lag distance, h, and the unit lag tolerance, h_{tol}, are scalar quantities.

2.1 Spatial relationship

Spatial relationship can be used to find out how neighboring values are related. There are different types of spatial correlation functions which define the spatial relationship. They are covariance, correlogram, traditional variogram, traditional cross variogram (in the case of two or more variables), general relative variogram, pairwise relative variogram, madogram, rodogram, and the traditional variogram for transformed variables, (e.g. variogram of logarithms of the variable, variogram for indicator variables and variogram for Gaussian variables). The definition for all of the spatial correlation functions will be presented in this chapter. Some of these spatial functions are named as robust variogram estimator (e.g. general and pairwise relative variogram). Among these spatial correlation functions, the traditional variogram is the most commonly used spatial relationship in geostatistics. In this study the focus is on the traditional variogram. The same methodology which will be presented in subsequent chapters can be used for other spatial correlation functions, but the traditional variogram function should be replaced with appropriate spatial correlation function.

In practice, several difficulties are encountered in estimating the variogram. These considerations are discussed below.

Decision of stationarity

The assumption of stationarity requires that the proposed geostatistical model based on our sampled data, can adequately describe the behaviour of the population. The goal is to infer the population based on the sample data. So, we should make an informed decision regarding what information we can use to describe the region of interest, this is called the decision of stationarity (Kelkar and Perez, 2002). In geostatistical study two kinds of stationarities can be defined, they are first order and second order stationarities. The first order stationarity is as below:

$$E\{Z(\mathbf{h})\} = E\{Z(\mathbf{u}+\mathbf{h})\} = m \quad \text{.....(2.1)}$$

It means that the expected value of a random variable at **u** is the same as the expected value of a random variable **h** lag distance away. Therefore first order stationarity means that the expected value across the region is the same. If we divide the region into small subregions and calculate the mean within each subregion then the means should be approximately the same in the case of first order stationarity (Kelkar and Perez, 2002). If the mean varies significantly from a subregion to another subregion, then there is a trend in the data. One of the most important parts of geostatistical modelling is to find the correct trend model if the data show a systematic trend. The trend function can be developed by a regression technique, inverse distance weighting and moving window averaging. This trend should be removed before variogram modelling and geostatistical simulation.

Second order stationarity uses the variance at each location, and it assumes that the variance is constant across the region. Therefore,

$$Var\{Z(\mathbf{h})\} = Var\{Z(\mathbf{u}+\mathbf{h})\} = \sigma^2 \quad \text{.....(2.2)}$$

By using the first and second order stationarities the relationship between the covariance and variogram can be obtained:

$$\gamma(\mathbf{h}) = \sigma^2 - C(\mathbf{h}) \quad \text{.....(2.3)}$$

where σ^2 is the variance of data, $\gamma(\mathbf{h})$ is the variogram and $C(\mathbf{h})$ is the covariance function.

Direction of continuity

Variogram and geological continuity are usually anisotropic, that is, direction dependent (Kelkar and Perez, 2002). Directions of continuity are most often known from geological interpretation or preliminary contouring of the data. At first stage of geostatistical study creating a kriged map can show directions of continuity for the spatial data set. Although, producing this kriged map requires a variogram itself but, it provides some idea about the large scale direction of continuity (Journel and Hujibregts, 2002).

Another common approach is to calculate and plot the variogram map. The variogram is calculated for a large number of directions and distances. Then the variogram values are posted on a map where the centre of the map is the lag distance of zero. But before plotting the variogram map we need to find the maximum lag distance and lag increment to calculate the variogram. The sampling pattern may suggest reasonable distance parameters (Deutsch, 2002).

The apparent anisotropy in the variogram can be checked further by calculating the lag distance at which the estimated variogram in each direction reaches the sill.

Robust estimation of variogram

In geostatistics, the variogram is the most commonly used statistical measure to describe spatial relationship. Variogram defines the connectivity of two points; it does not allow simultaneous definition of connectivity of multiple points, which may be important in certain conditions.

The definition of the variogram is based on the difference in a variable measured at two points located a certain distance apart. This traditional variance based definition is convenient because it allows us to define the uncertainty with respect to estimation at unsampled location. If we quantify the uncertainty using some other method, we may not need to use a variance based spatial relationship, such as the variogram. Once the restriction of a variance based spatial relationship is removed, the traditional semivariogram equation can be modified as (Kelkar and Perez, 2002):

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left\{ \left| Z(\mathbf{h}) - Z(\mathbf{u}+\mathbf{h}) \right|^{w} \right\} \quad \text{where} \quad 0 < w \leq 2 \qquad \text{.....(2.4)}$$

If we choose w=2, the above formula becomes the traditional variogram. If w=1, it is called the madogram, which is a measure of the absolute deviation between the two values. If w=0.5, it is called a rodogram, which is the square root of absolute deviation. The smaller the value of w, the more resistant the spatial relationship is to outlier data. Outlier data are data points that fall outside the norm. For a normal distribution, a data point falling outside the mean plus or minus three times of standard deviation can be considered as outlier.

Modified variograms in the case of biased sampling

We can modify the traditional variogram in order to remove the impact of biased sampling on the estimation of variogram. For example in reservoir characterization, we may deal with biased data because our first few wells may be drilled based on limited information. As information is gathered from these wells, next wells are drilled on the basis of the additional information. It is obvious that we drill the wells in areas where the potential of oil recovery is maximum. In the case of biased sampling, in order to capture the direction of continuity and anisotropy better we can use modified variograms. They are general relative variogram, pairwise relative variogram, and non-ergodic variogram, covariance and correlogram.

The definitions for these modified variograms are written below:

- *General relative variogram*

  The general relative variogram is defined as

$$\gamma_{GR}(\mathbf{h}) = \frac{\gamma(\mathbf{h})}{\left(\frac{m_{-\mathbf{h}} + m_{+\mathbf{h}}}{2} \right)^2} \qquad \text{.....(2.5)}$$

  Where $\gamma(\mathbf{h})$ is the traditional variogram and $m_{-\mathbf{h}}$ is the local mean at tail value and $m_{+\mathbf{h}}$ is the local mean at head value. Normalizing the variogram with a local

mean eliminates the influence of the variations in lag mean (Kelkar and Perez, 2002).

- *Pairwise relative variogram*

  The pairwise relative variogram is defined as

$$\gamma_{PR}(\mathbf{h}) = \frac{2}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \left[\frac{z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})}{z(\mathbf{u}_i) + z(\mathbf{u}_i + \mathbf{h})} \right]^2 \quad \text{.....(2.6)}$$

  The pairwise relative variogram normalizes the traditional variogram the same as general relative variogram. The only difference between these two method of normalization is that, in the pairwise relative variogram, each pair difference squared is normalized with respect to the square of the pair mean but in general relative variogram each pair difference squared is normalized with respect to the square of the lag mean (Kelkar and Perez, 2002).

- *Non-ergodic variogram, covariance and correlogram*

  A random function is called ergodic if its mean or covariance or variogram coincide with the corresponding spatial averages calculated over the single available realization which is the sample value at each location. If we have non-ergodic behaviour in our data we should modify the definition for variogram (Deutsch, 2002).

  - *Non-ergodic covariance*:

$$C_{NE}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} z(\mathbf{u}_i) z(\mathbf{u}_i + \mathbf{h}) - m_{-\mathbf{h}} m_{+\mathbf{h}} \quad \text{.....(2.7)}$$

  - *Non-ergodic variogram*:

$$\gamma_{NE}(\mathbf{h}) = C(\mathbf{0}) - C_{NE}(\mathbf{h}) \quad \text{.....(2.8)}$$

  - *Non-ergodic correlogram*:

$$\rho_{NE}(\mathbf{h}) = \frac{C_{NE}(\mathbf{h})}{\sigma_{-\mathbf{h}} \sigma_{+\mathbf{h}}} \quad \text{.....(2.9)}$$

Modified Variograms in the case of Outlier Data

Outlier data can significantly affect variogram estimation. Use of an extreme value in variogram estimation can amplify the effect because the squared difference between a data pair is used. If the difference between a given pair is several orders of magnitude, the squared difference is large enough to influence the estimated variogram at a particular lag distance. This may create instability in the variogram estimation and also may prevent us from clearly identifying the spatial structure.

We have two options for dealing with the outlier data. The first option is to remove the outlier data point from the estimation process. While this is the simplest option, it is reasonable only if we have a physical reason for its omission. The second option is to use some type of nonlinear transformation to minimize the variation. These include the log, power, rank, indicator and normal score transforms (Kelkar and Perez, 2002).

- *Log transform*

  The most commonly used transform is to use the natural logarithm of the sample value. Log transform is simple to use but may create difficulties in estimating value at unsampled locations. Our semivariogram formula then becomes:

$$\gamma_L(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \left[\log\left[z(\mathbf{u}_i)\right] - \log\left[z(\mathbf{u}_i + \mathbf{h})\right] \right]^2 \quad \text{.....(2.10)}$$

- *Power transform*

  In this transformation the original value, $z(\mathbf{u}_i)$, is replaced by $\left(z(\mathbf{u}_i)\right)^p$. Therefore the corresponding variogram is:

$$\gamma_{PT}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \left[\left(z(\mathbf{u}_i)\right)^p - \left(z(\mathbf{u}_i + \mathbf{h})\right)^p \right]^2 \quad \text{..... (2.11)}$$

  where p is a positive number and less than 1. The smaller the power we use, the smaller the variation of the data values. The most commonly used value of p is 0.5, which represents the square root of the sample data value.

- *Rank transform*

  To calculate the ranking of a given sample value, all the sample data are arranged in an ascending order. The rank of the i^{th} value in the sequence is calculated by

  $$R_i = \frac{i}{n+1} \quad \text{.....(2.12)}$$

  Where n is the total number of sample points and R_i is the rank of the i^{th} value in sequence. Using the rank transform ensures that no significance difference exists between sample values because all the sample values fall between zero and one.

- *Indicator transform*

  The indicator transform allows transformation of a continuous variable into a discrete variable. Indicator transformation is defined as below:

  $$i(\mathbf{u}, z_k) = \begin{cases} 1 & if \quad Z(\mathbf{u}) \le z_k \\ 0 & if \quad Z(\mathbf{u}) > z_k \end{cases} \quad \text{.....(2.13)}$$

  By specifying multiple threshold values, we can define multiple indicator values at each threshold. Defining each sample point in terms of either zero or one eliminates the effect of outlier data. One disadvantage of defining the indicator values is that the exact differences between the data values in a particular class are lost. Defining additional thresholds can approximately remove this disadvantage. In addition to removing the outlier data, indicator transformation has two more advantages. First, by appropriately defining the thresholds values and estimating the indicator variogram at each threshold, we can examine how the sample values are connected at different thresholds. For example, we may observe that the low values exhibit better continuity than high values. Another advantage for indicator transform is that we can transform our qualitative data (rock type) to quantitative data (0 and 1).

- *Normal score transform*

  The normal score transform allows transformation of sample data into equivalent data that follow a normal distribution with a mean of zero and the variance of 1. The normal score transformation is:

  $$x_{NS} = G^{-1}\left(F(x)\right) \quad \text{.....(2.15)}$$

  Where $G^{-1}(\bullet)$ is the inverse of the standard Gaussian cumulative distribution function and $F(x)$ is the global cumulative distribution function of the sample data.

  After this transformation the normal score values can be used, x_{NS}, instead of the original values to calculate the semivariogram. With this type of transformation we can remove the outlier effect in our data. Another advantage for normal score transformation is that the certain estimation techniques work better with normal score transformed data. After obtaining the estimation with the transformed data, we can back-transform the data to original variable values.

2.2 Calculating and fitting the experimental variogram

After getting an idea about the direction of continuity in our data set, the experimental variograms in two main directions of continuity are calculated. This calculation needs reasonable tolerance parameters. These tolerances are defined in this section. They can be obtained by some logical decisions, but in the next chapters a methodology will be presented to optimize these tolerance parameters.

In order to use variograms for estimation and simulation, we need a model which represents the spatial variability. Common practice consists of fitting experimental variograms with a nested combination of proven models such as the spherical, exponential, and Gaussian models. Both hand fitting and semi-automatic fitting (by using `varfit`; Neufeld and Deutsch, 2004) can be used to model the variogram.

Variogram tolerance parameters

Choosing the tolerance parameters in the case of irregularly gridded data is complicated. As mentioned before, the goal of this study is to find a smart way to optimize these tolerance parameters in the case of irregularly spaced data. The tolerance parameters should be chosen in such a way that there is enough number of pairs for a reliable variogram value. Usually in real cases there are more data in vertical direction than horizontal direction; therefore it is better to separate these directions. The tolerance parameters are different in these two cases.

Tolerances in vertical directions

There are four tolerance parameters in this case; they are (Deutsch, 2002):

- Unit lag separation distance, h (h is a scalar quantity); it is equal to data spacing in the case of regularly spaced data. In petroleum applications, the vertical data are regularly spaced, therefore choosing the unit lag separation distance is not a problem.
- The distance tolerance, h_{tol} (similar to h, h_{tol} is a scalar quantity); in real case this parameter usually takes the values of $0.25h$, $0.5h$ and $0.75h$. The value of $0.5h$ is used in almost all of the cases. The value of $0.25h$ is used when there are many data on a nearly regular grid and in the case of small number of pairs for each lag, it is recommended to use $0.75h$. In next chapter this parameter will be optimized to have a value between 0 and h.
- The angle tolerance, a_{tol}; this parameter is used when there is some deviation from vertical direction.
- A bandwidth parameter, b; this parameter is defined to limit the region for finding data pairs for variogram calculation in vertical direction. After a certain lag distances bandwidth parameter is applied to limit the tolerance region.

Tolerances in horizontal directions

There are six tolerance parameters in this case; they are (Deutsch, 2002):

- Unit lag separation distance, h; the properties of this parameter are the same as vertical case.
- The distance tolerance, h_{tol}; the properties of this parameter are the same as vertical case.
- The horizontal angle tolerance, a^h_{tol}; this parameter can be applied to limit the direction which we are interested to calculate the associated variogram. In the case of omnidirectionality, it can be set to 90 degree or a greater angle to have all of the points in horizontal plane. The omnidirectional experimental variogram averages the variability over all directions.
- The horizontal bandwidth, b_{hor}; similar to the vertical bandwidth, it is used to define a limited tolerance region in horizontal plane. In the case of omnidirectionality, the horizontal bandwidth should be set to a large number to have all of the points in all directions to calculate the omnidirectional variogram value.
- The vertical angle tolerance, a^v_{tol}, used to define an angle tolerance about the horizontal plane. It should be set to a small value because of the large variability in vertical direction which can affect the true variogram calculation for a specific stratigraphic layer.
- The vertical bandwidth, b_{ver}; this bandwidth relates to the vertical angle tolerance and is defined to limit the stratigraphic layer. It should be set to a small value to have a good approximation of the variogram.

For both vertical and horizontal lag distances, the number of distance lags should be chosen in such a way that the total distance used for variogram calculation is about one half of the field size.

Remarks

The choice of variogram model has a major affect on kriging and kriging-based simulation. All the methods for modelling the spatial variability have some advantages and may screen artifacts from sparse data, non-stationarity, outliers, etc.

Chapter 3

Improving Variogram Calculations

3.1 Methodology and principles

The GSLIB (Deutsch and Journel, 1998) program, gamv, can be used to calculate the experimental variogram points for a given irregular data set. The user should define some tolerance parameters (see section 2.2) based on the data locations to calculate the experimental points. The parameter file for gamv is written below:

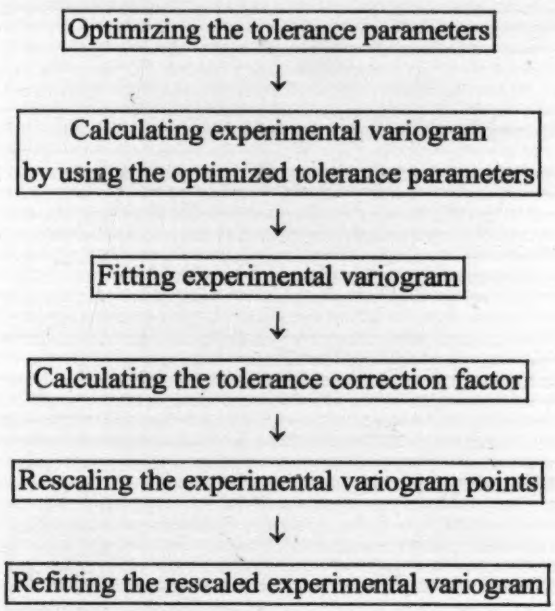
```
                  Parameters for GAMV
                  *******************

START OF PARAMETERS:
data.dat                                -file with data
1   2   0                               -   columns for X, Y, Z coordinates
1   3                                   -   number of variables,col numbers
-1.0e21     1.0e21                      -   trimming limits
gamv.out                                -file for variogram output
10                                      -number of lags
10                                      -lag separation distance
5                                       -lag tolerance
1                                       -number of directions
0.0  90.0 400.0   0.0  90.0  400.0      -azm,atol,bandh,dip,dtol,bandv
0                                       -standardize sills? (0=no, 1=yes)
1                                       -number of variograms
1   1   1                               -tail var., head var.,variogram type
```

The bolded texts are the parameters which are used to define the tolerance parameters. They should be changed properly in the case of 2D or 3D data set to get a reasonable and

meaningful variogram. By changing each of these parameters the result would be different. There should exist a set of tolerance parameters that can give a reasonable result close to the true variogram. It means that after fitting the experimental points the variogram model should have minimum error when comparing to the true variogram. This minimization should be carried out for different sets of tolerance parameters.

As was explained in Chapter 1 the goal in this study is to introduce a methodology to improve variogram calculation and fitting. By improving we mean that the final fitted variogram has less error for the optimized tolerance parameters compared to the true variogram which is almost always unknown. The main algorithm for this improvement is illustrated below; the algorithm has 6 steps to update the experimental variogram points.



Each step is discussed briefly in next section.

3.2 Implementation aspects

3.2.1 Optimizing tolerance parameters

To optimize the tolerance parameters a penalty function is defined and the optimal tolerance parameters are obtained by minimizing this penalty function. This penalty function gives a penalty for each set of parameters, this can be obtained by adding all of the errors (between the fitted variogram and the true or assumed to be true variogram) associated to each lag distance (and for each set of parameters). Figure 3.1 shows schematically the difference function between the fitted variogram and the reference variogram which is used in the definition of the penalty function.

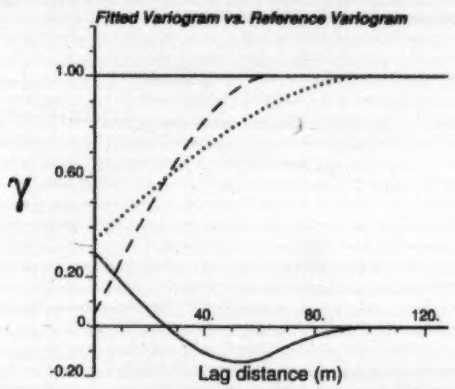


Figure 3.1. Fitted variogram (dotted line) versus the reference variogram (dashed line) and their difference (fitted minus reference; solid line)

So the error can be written in integral form as below:

$$p_{i_{realz}}(\theta) = \frac{1}{2a}\int_0^{2a}\left[\gamma^{fit}_{i_{realz}}(\mathbf{h};\theta) - \gamma^{ref}(\mathbf{h})\right]^2 \cdot wt(\mathbf{h}) \cdot d\mathbf{h} \quad \text{.....(3.1)}$$

Where:

$p_{i_{realz}}(\theta)$; Penalty function for the i_{realz} realization and set of tolerance parameters θ

i_{realz} ; Realization number

θ ; A set of tolerance parameters

a ; Reference variogram range

$\gamma^{fit}_{i_{tol}}(\mathbf{h};\theta)$; Fitted variogram of the experimental variogram points based on the tolerance parameters θ.

$\gamma^{ref}(\mathbf{h})$; Reference variogram

$wt(\mathbf{h})$; Inverse distance weight associated to each lag distance

The inverse distance weighting associated to each lag is calculated from below formula:

$$wt(\mathbf{h}) = \begin{cases} \dfrac{1}{\mathbf{h}+\varepsilon} & ; \quad \textit{for } \mathbf{h} \textit{ such that } \gamma^{fit}(\mathbf{h};\theta) > \gamma^{ref}(\mathbf{h}) \\ \dfrac{0.5}{\mathbf{h}+\varepsilon} & ; \quad \textit{for } \mathbf{h} \textit{ such that } \gamma^{fit}(\mathbf{h};\theta) < \gamma^{ref}(\mathbf{h}) \end{cases} \qquad \text{.....(3.2)}$$

The reason for inverse weighting of the difference between the true variogram and the reference variogram is that the small lag distances which show the short variability get more weight and the large lag distances get less weight. The small number, ε, is used in the denominator because in this case the weight at $\mathbf{h}$=0 can be defined (which is a large number). θ has different number of components depending on 2D or 3D cases. In 2D case θ has at most 4 components: lag separation distance, h, lag tolerance, h_{tol}, angle tolerance, a_{tol}, and the bandwidth, b. In the case of omnidirectionality where the angle tolerance is 90 degrees and the bandwidth is a large value, θ has 2 components of lag separation distance and the lag tolerance only. These four parameters are shown in Figure 3.2.

In 3D case since we are dealing with vertical and horizontal (minor and major) directions, therefore two different sets of parameters can be defined. For the vertical lag distance (Figure 3.3.) θ has at most 4 components these are the unit lag separation distance, h, lag tolerance, h_{tol}, angle tolerance, a_{tol}, and the bandwidth, b.

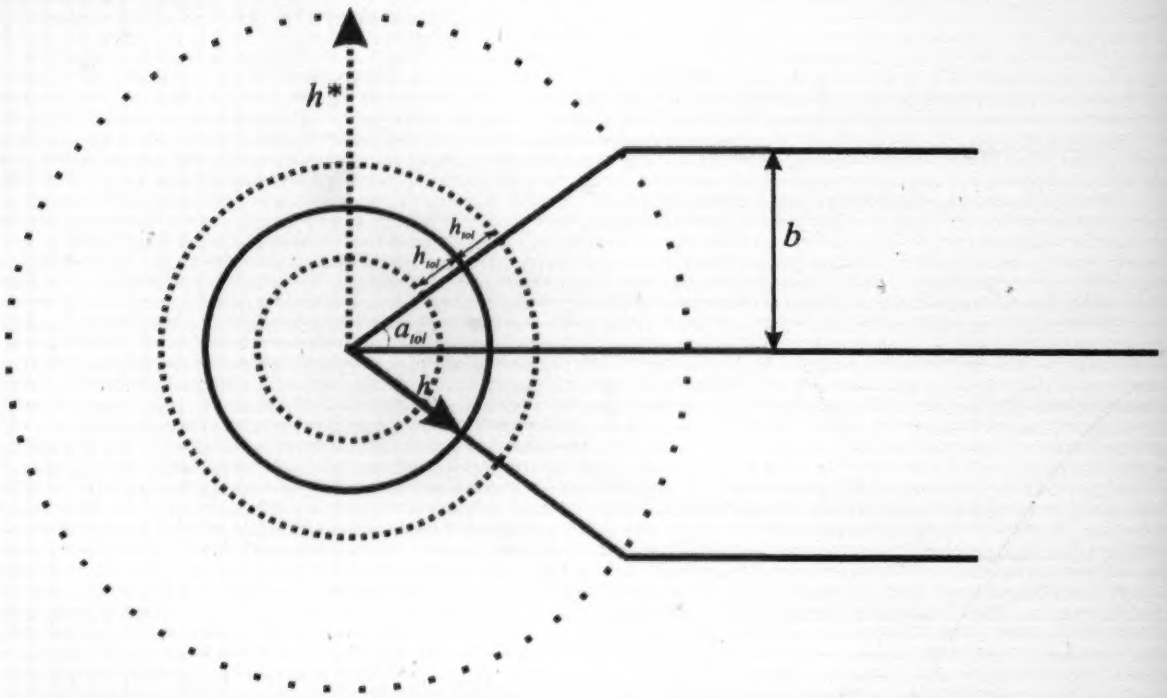


Figure 3.2. Lag separation distance, h, lag tolerance, h_{tol}, azimuth tolerance, a_{tol} and bandwidth, b

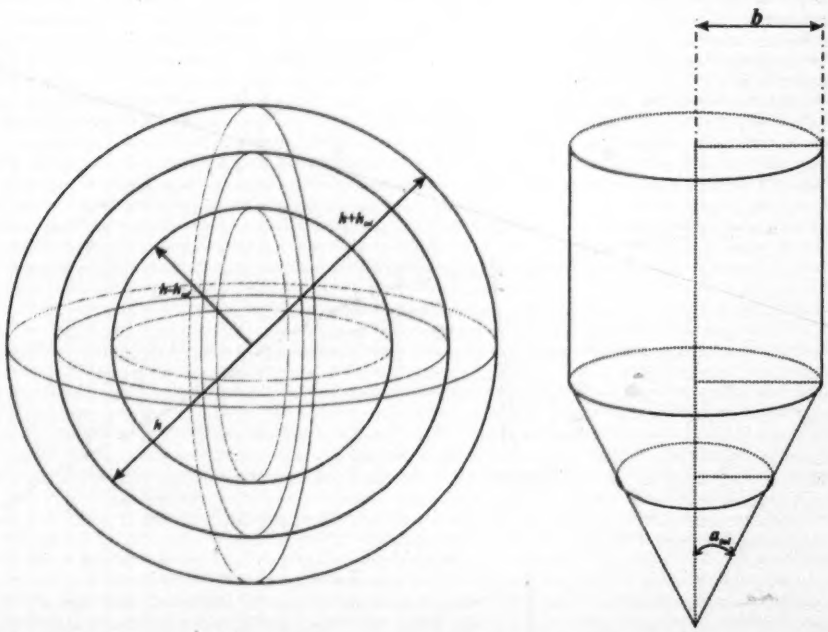


Figure 3.3. Lag separation distance, h, lag tolerance, h_{tol}, angle tolerance, a_{tol} and bandwidth, b in three dimensional case-vertical lag distance

For the horizontal lag distance (Figure 3.4) θ has at most 6 components: unit lag separation lag distance, h, the lag tolerance, h_{tol}, the vertical angle tolerance, a^{v}_{tol}, the horizontal angle tolerance, a^{h}_{tol}, the vertical bandwidth, b_{ver} and the horizontal bandwidth, b_{hor}. The number of components can be decreased in case of omnidirectionality or in the case with no bandwidth.

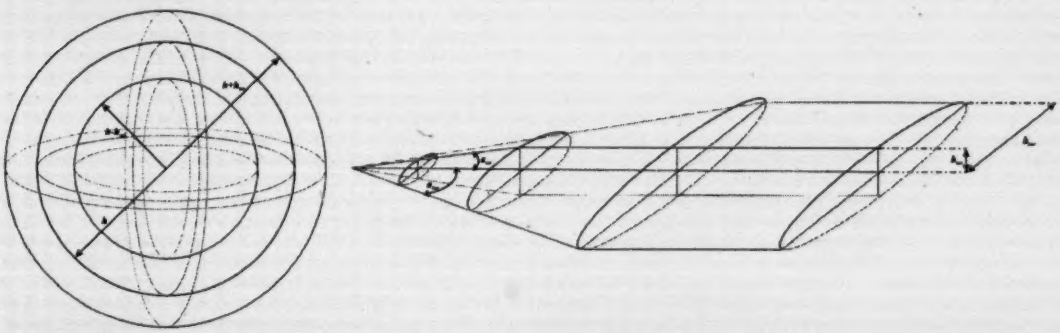


Figure 3.4. Lag separation distance, h; lag tolerance, h_{tol}; vertical angle tolerance, a^{v}_{tol}; horizontal angle tolerance, a^{h}_{tol}; vertical bandwidth, b_{ver} and horizontal bandwidth, b_{hor} in three dimensional case-horizontal lag distance

Besides the tolerance parameters another parameter can be defined in order to capture the relationship between the tolerance parameters and the penalty function. This parameter which is named tolerance ratio is a function of all of the tolerance parameters which are used to calculate the experimental variogram. A tolerance ratio is equal to the tolerance area (in 2D case) or volume (in 3D case) divided by the area or the volume of the semi-circle or hemi-sphere with a radius of $2h$ (h is the unit lag distance). For example this tolerance ratio is 100 % in the case of the calculation of the omnidirectional variogram (the only tolerance parameters are unit lag distance and the lag tolerance) with the lag tolerance, h_{tol}, which is equal to unit separation lag distance; h. Figure 3.5 schematically explains the tolerance ratio in 2D case. The tolerance ratio will be used as one of the parameters for optimizing the tolerance parameters. In general there are two different cases for calculating the tolerance ratio; 2D (omnidirectional and general case) and 3D (omnidirectional, horizontal and vertical lag distances). These cases are shown in Figures 3.2, 3.3 and 3.4.

In next sections the formulas for tolerance ratio are given for different cases. The proofs for 3D are presented in Appendix A.

3.2.1.1 Tolerance ratio in 2D

Omnidirectional case

In this case tolerance ratio depends only on unit lag distance, h and the lag tolerance, h_{tol}. The bandwidth and the angle tolerance should be set to a large number and 90 degree in order to calculate the omnidirectional experimental variogram by using gamv program in GSLIB. The area function $A(r)$ is the area of the semi-circle with the radius of r. (see Figure 3.5)

$$A(r)=\frac{\pi}{2}r^2 \qquad \text{.....(3.3)}$$

Therefore the tolerance ratio is:

$$Tol=\frac{A(h+h_{tol})-A(h-h_{tol})}{A(2h)}=\frac{h_{tol}}{h} \qquad \text{.....(3.4)}$$

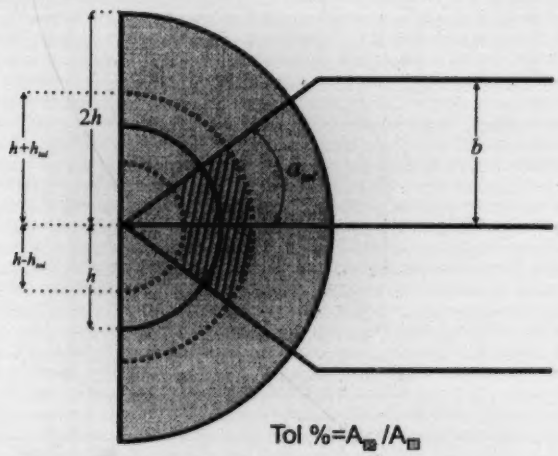


Figure 3.5. Tolerance ratio for two dimensional case

General case

In this case the tolerance area is function of unit lag distance, h, the lag tolerance, h_{tol}, angle tolerance, a_{tol} and the bandwidth, b. Depending on the magnitude of the unit lag distance, h, three cases would be happen. To specify these three cases another parameter, h^*, can be defined which is function of the angle tolerance, a_{tol} and the bandwidth, b. h^* can be calculated by using below formula (see Figure 3.2):

$$h^* = \frac{b}{\sin a_{tol}} \quad \text{.....(3.5)}$$

It can be seen from Figure 3.2 that when $h \le \left(h^* - h_{tol}\right)$, the tolerance area is made by using the unit lag distance, lag tolerance and angle tolerance (note that the tolerance area in this case is independent of the magnitude of bandwidth). If $h > \left(h^* + h_{tol}\right)$ the tolerance area is function of unit lag distance, lag tolerance and bandwidth (in this case the tolerance are is independent of the magnitude of angle tolerance) and if $\left(h^* - h_{tol}\right) < h \le \left(h^* + h_{tol}\right)$ then the tolerance area is function of all four tolerance parameters.

The three cases for different lag distances can be defined as:

1. $h \le \left(h^* - h_{tol}\right)$;

$$A_1(r,\alpha) = \alpha \cdot r^2 \quad \text{.....(3.6)}$$

$$Tol = \frac{A_1\left(h + h_{tol}, a_{tol}\right) - A_1\left(h - h_{tol}, a_{tol}\right)}{A(2h)} = \left(\frac{2a_{tol}}{\pi}\right) \cdot \left(\frac{h_{tol}}{h}\right) \quad \text{.....(3.7)}$$

a_{tol} is in radians.

2. $h > \left(h^* + h_{tol}\right)$;

$$A_2(r,b) = \left[\left(\frac{b}{r}\right) \cdot \sqrt{1 - \left(\frac{b}{r}\right)^2} + \arcsin\left(\frac{b}{r}\right)\right] \cdot r^2 \quad \text{.....(3.8)}$$

$$Tol = \frac{A_2(h+h_{tol},b) - A_2(h-h_{tol},b)}{A(2h)} \quad \text{.....(3.9)}$$

3: $(h^* - h_{tol}) < h \le (h^* + h_{tol})$;

$$Tol = \frac{A_1(h^*,a_{tol}) - A_1(h-h_{tol},a_{tol}) + A_2(h+h_{tol},b) - A_2(h^*,b)}{A(2h)} \quad \text{.....(3.10)}$$

3.2.1.2 Tolerance ratio in 3D

Omnidirectional case

Similar to the 2D case, the tolerance ratio only depends on the unit lag distance, h, and the lag tolerance, h_{tol}, in this case instead of area function; the volume function should be defined which is equal to the volume of the hemisphere with a radius of r.

$$V(r) = \frac{2\pi}{3} \cdot r^3 \quad \text{.....(3.11)}$$

Therefore from equation 3.11, we will have:

$$Tol = \frac{V(h+h_{tol}) - V(h-h_{tol})}{V(2h)} = \frac{1}{4} \cdot \left(\frac{h_{tol}}{h}\right) \cdot \left[3 + \left(\frac{h_{tol}}{h}\right)^2\right] \quad \text{.....(3.12)}$$

General case

For the general 3D case, two scenarios can be considered; the first one is for the vertical lag distance (Figure 3.3) and the second one is for horizontal lag distance (Figure 3.4).

a) Vertical lag distance

In this case the tolerance volume is a function of unit lag distance, h, the lag tolerance, h_{tol}, angle tolerance, a_{tol} and the bandwidth, b (see Figure 3.3). Based on the magnitude of the unit lag distance three different cases exist.

1. $h \le (h^* - h_{tol})$;

$$V_1(r,\alpha) = \frac{2\pi}{3} \cdot (1 - \cos\alpha) \cdot r^3 \quad \text{.....(3.13)}$$

$$Tol = \frac{V_1(h+h_{tol}, a_{tol}) - V_1(h-h_{tol}, a_{tol})}{V(2h)} \quad \text{.....(3.14)}$$

Therefore from equations 3.11, 3.13 and 3.14, we will have:

$$Tol = \frac{1}{4}\cdot(1-\cos a_{tol})\cdot\left(\frac{h_{tol}}{h}\right)\cdot\left[3+\left(\frac{h_{tol}}{h}\right)^2\right] \quad \text{.....(3.15)}$$

2. $h > (h^* + h_{tol})$;

$$V_2(r,b) = \frac{2\pi}{3}\cdot\left[1-\left(1-\left(\frac{b}{r}\right)^2\right)^{\frac{3}{2}}\right]\cdot r^3 \quad \text{.....(3.16)}$$

$$Tol = \frac{V_2(h+h_{tol}, b) - V_2(h-h_{tol}, b)}{V(2h)} \quad \text{.....(3.17)}$$

3. $(h^* - h_{tol}) < h \le (h^* + h_{tol})$;

$$Tol = \frac{V_1(h^*, a_{tol}) - V_1(h-h_{tol}, a_{tol}) + V_2(h+h_{tol}, b) - V_2(h^*, b)}{V(2h)} \quad \text{.....(3.18)}$$

b) Horizontal lag distance

In this case the tolerance volume is a function of the unit lag distance, h, the lag tolerance, h_{tol}, the vertical angle tolerance, a^v_{tol}, the horizontal angle tolerance, a^h_{tol}, the vertical bandwidth, b_{ver} and the horizontal bandwidth, b_{hor} (see Figure 3.4):

$$Tol = f(h, h_{tol}, a^v_{tol}, a^h_{tol}, b_{ver}, b_{hor}) \quad \text{.....(3.19)}$$

As before the magnitude of the unit lag distance results in different scenarios:

1. $h \le (h^*_{ver} - h_{tol})$; $(h^*_{ver} = \frac{b_{ver}}{\sin a^v_{tol}})$

$$V_3(r,\alpha,\beta)=\frac{2\pi}{3}\cdot\left[1-I_1(\alpha,\beta)\right]\cdot r^3 \quad \text{.....(3.20)}$$

Where

$$I_1(\alpha,\beta)=\frac{1}{2\pi}\cdot\int_0^{2\pi}\left(\frac{\cot^2\alpha\cdot\cos^2\theta+\cot^2\beta\cdot\sin^2\theta}{1+\cot^2\alpha\cdot\cos^2\theta+\cot^2\beta\cdot\sin^2\theta}\right)^{\frac{1}{2}}\cdot d\theta \quad \text{.....(3.21)}$$

$$Tol=\frac{V_3\left(h+h_{tol},a_{tol}^h,a_{tol}^v\right)-V_3\left(h-h_{tol},a_{tol}^h,a_{tol}^v\right)}{V(2h)} \quad \text{.....(3.22)}$$

Therefore from equations 3.20, 3.21 and 3.22, we will have:

$$Tol=\frac{1}{4}\cdot\left[1-I_1\left(a_{tol}^h,a_{tol}^v\right)\right]\cdot\left(\frac{h_{tol}}{h}\right)\cdot\left[3+\left(\frac{h_{tol}}{h}\right)^2\right] \quad \text{.....(3.23)}$$

2. $\left(h_{ver}^*+h_{tol}\right)<h\le\left(h_{hor}^*-h_{tol}\right)$;

$$V_4(r,b,\alpha)=\frac{2\pi}{3}\cdot\left[1-I_2(r,b,\alpha)-\frac{3\sin(2\alpha)}{8}\cdot\left(\frac{b}{r}\right)\right]r^3 \quad \text{.....(3.24)}$$

Where

$$I_2(r,b,\alpha)=\frac{1}{2\pi}\cdot\int_0^{2\pi}\left[\frac{\csc^2\alpha\cdot\cos^2\theta+\left(\frac{r}{b}\right)^2\cdot\sin^2\theta-1}{\csc^2\alpha\cdot\cos^2\theta+\left(\frac{r}{b}\right)^2\cdot\sin^2\theta}\right]^{\frac{3}{2}}\cdot d\theta \quad \text{.....(3.25)}$$

$$Tol=\frac{V_4\left(h+h_{tol},b_{ver},a_{tol}^h\right)-V_4\left(h-h_{tol},b_{ver},a_{tol}^h\right)}{V(2h)} \quad \text{.....(3.26)}$$

3. $\left(h_{ver}^*-h_{tol}\right)<h\le\left(h_{ver}^*+h_{tol}\right)$;

$$Tol=\frac{V_3\left(h_{ver}^*,a_{tol}^h,a_{tol}^v\right)-V_3\left(h-h_{tol},a_{tol}^h,a_{tol}^v\right)+V_4\left(h+h_{tol},b_{ver},a_{tol}^h\right)-V_4\left(h_{ver}^*,b_{ver},a_{tol}^h\right)}{V(2h)}$$

.....(3.27)

4. $h > \left(h^{*}_{hor} + h_{tol}\right)$;

$$V_5(r,a,b) = \frac{2\pi}{3} \cdot \left[1 - I_3(r,a,b)\right] \cdot r^3 \quad \text{.....,,....(3.28)}$$

Where

$$I_3(r,a,b) = \frac{1}{2\pi} \cdot \int_0^{2\pi} \left[\frac{\left(\frac{r}{a}\right)^2 \cdot \cos^2\theta + \left(\frac{r}{b}\right)^2 \cdot \sin^2\theta - 1}{\left(\frac{r}{a}\right)^2 \cdot \cos^2\theta + \left(\frac{r}{b}\right)^2 \cdot \sin^2\theta} \right]^{\frac{3}{2}} d\theta \quad \text{.....(3.29)}$$

$$Tol = \frac{V_5\left(h + h_{tol}, b_{hor}, b_{ver}\right) - V_5\left(h - h_{tol}, b_{hor}, b_{ver}\right)}{V(2h)} \quad \text{.....(3.30)}$$

5. $\left(h^{*}_{hor} - h_{tol}\right) < h \le \left(h^{*}_{hor} + h_{tol}\right)$;

$$Tol = \frac{V_4\left(h^{*}_{hor}, b_{ver}, a^{h}_{tol}\right) - V_4\left(h - h_{tol}, b_{ver}, a^{h}_{tol}\right) + V_5\left(h + h_{tol}, b_{hor}, b_{ver}\right) - V_5\left(h^{*}_{hor}, b_{hor}, b_{ver}\right)}{V(2h)}$$

.....(3.31)

3.2.1.3 Methodology

The whole procedure for optimizing the tolerance parameters is as below:

1. Use spatial bootstrap (Deutsch, 2004) to obtain L values (realizations) for each of the data locations.
2. Determine different reasonable sets of tolerance parameters for variogram calculation.
3. For each set of tolerance parameters and each realization:
   a. Calculate the experimental variogram.
   b. Fit the experimental variogram with a variogram model, $\gamma^{fit}_{l_{reals}}(\mathbf{h};\theta)$, by using `varfit` program (Neufeld and Deutsch, 2004).
   c. Calculate the penalty function, $p_{l_{reals}}(\theta)$, by using equation 3.1.

4. For each set of tolerance parameters, calculate final penalty by averaging over different calculated penalty for different spatial bootstrap realizations.
5. Map penalty function as a function of tolerance parameters.
6. Find the optimum tolerance parameters by minimizing the penalty function.

Spatial bootstrap

Spatial bootstrap (Deutsch, 2004) is used to assess and quantify the uncertainty in the variogram at each lag. The spatial bootstrap procedure is as follow:

1. Preliminary analysis:
    a. Assemble the representative distribution of the random variable Z, $F(z)$;
    b. Define a 3-D variogram model $\gamma(h)$ of the normal scores of the random variable Z;
    c. Decompose the n by n covariance matrix by the Cholesky decomposition into product of upper (U) and lower (L) triangular matrices: $C = LU$
2. Generate a new set of data, z, as

$$z = F_z^{-1}\left(G\left(Lw\right)\right) \quad \text{.....(3.32)}$$

    where w is a n by 1 vector of independent Gaussian values and $G(\cdot)$ denotes the standard Gaussian cumulative distribution function.
3. Calculate the statistic of interest from the resampled dataset.
4. Repeat Steps 2 and 3 many times, say, L=100.
5. Establish the distribution of uncertainty in the calculated statistic.

This uncertainty has effect on the selection of the optimal tolerance parameters. The penalty function, $p_{l_{real}}(\theta)$, can be calculated for different realizations which are output of spatial bootstrap. Spatial bootstrap needs the original data and also the reference variogram to create different realizations. This reference variogram is the same as $\gamma^{ref}(\mathbf{h})$ which is used to calculate the penalty function, $p_{l_{real}}(\theta)$. Basically to obtain the

tolerance parameters for a given data set first the spatial bootstrap is performed to get different realization at each data location. Two assumptions are made before performing spatial bootstrap; the first one is that the initial data distribution is representative of the entire population and the second one is that the data are spatially correlated; this spatial correlation is represented by the reference variogram which is unknown and also is not modeled at the first step of the tolerance parameters calculations. In chapter 4, a sensitivity analysis will be shown for a real data set on how to choose reasonable variogram range and nugget effect for a single spherical structure for the reference variogram model. The reasonable assumptions for the variogram model could be single spherical structure with range of one-third of the variogram range. Declustering techniques might be performed to get the weights for the data points. Spatial bootstrap needs the distribution of the original data. The cumulative distribution function of the data is required for randomly drawing the data values for the bootstrap. After determining the reference distribution and variogram the spatial bootstrap can be applied to get L realizations for each of the data locations.

Experimental variogram calculations for different tolerance parameters

The GSLIB program, gamv, can be used to calculate the experimental variogram for different tolerance parameters. Reasonable range for tolerance parameters should be defined for calculations. This range should contain the optimal tolerance parameters. The range for unit lag distance or lag separation distance can be obtained by building the cumulative distribution function of the minimum distance between the data locations. There is just one value for the minimum distance which corresponds to a data location. The minimum distance between the data points is calculated as:

$$d_{i_{data}} = \min\left\{\left|\mathbf{u}_{i_{data}} - \mathbf{u}_{j_{data}}\right|\right\} \quad ; \quad i_{data} \neq j_{data} \text{ and } i_{data}, j_{data} = 1,2,3,\ldots,n_{data} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.33)$$

where $d_{i_{data}}$ is the minimum distance between the data points which corresponds to each data location, $\left|\mathbf{u}_{i_{data}} - \mathbf{u}_{j_{data}}\right|$ is the norm of the separation vector between data locations i_{data} and j_{data}. The cumulative distribution function for $d_{i_{data}}$ can be built and the reasonable

range of values for unit lag separation distance can be obtained by using this distribution. The values between $P10$ and $P90$ of the $d_{i_{data}}$ values are used for different lag separation distances. Because of the CPU cost, the number of unit lag distances for different variogram calculations, n_h, should not be large. To cover the whole range and test different lag separation distance, the following relationship is considered for calculating the unit lag distance for different cases.

$$h = \left(\frac{P90 - P10}{n_h} \right) \cdot i_h + P10 \quad ; \quad i_h = 0,1,2,\ldots,n_h \qquad \text{.....(3.34)}$$

Where h is the unit lag distance, $P10$ and $P90$ are the distances calculated from distribution of $d_{i_{data}}$. n_h is the number of different unit lag distance used for calculating the optimal one and i_h is the index for n_h.

For the range of the lag tolerance, the values between 0 % and 100 % of the unit lag distance are used. Therefore

$$h_{tol} = \left(\frac{i_{tol}}{n_{tol}} \right) \cdot h \quad ; \quad i_{tol} = 1,2,\ldots,n_{tol} \qquad \text{.....(3.35)}$$

where n_{tol} is the number of different lag tolerances used for calculating the optimal one. Similar to n_h, n_{tol} should not be a large number for CPU efficiency.

For azimuth tolerance and dip tolerance, we can use values between 0 and 90 degrees. The increments can be defined to calculate different angle tolerances. Again the number of increments has impact on CPU time. The range for horizontal and vertical bandwidth can be estimated by using some geological information. In the case of lack of geological information the minimum and maximum possible values for the bandwidths can be considered. The maximum possible value for horizontal bandwidth can be the maximum areal dimension of the field and for the vertical bandwidth it is the thickness of the stratigraphic layer. After obtaining reasonable ranges for the tolerance parameters, the experimental variogram is calculated for each set of tolerance parameters and for each realization obtained from spatial bootstrap.

Fitting the experimental variogram for each realization

In this step the compatible GSLIB program, `varfit` (Neufeld and Deutsch, 2004), is used to fit the experimental variogram points with an optimized variogram model, the experimental points are different for different tolerance parameters and different realizations. The fitted model, $\gamma^{fit}_{i_{realz}}(\mathbf{h};\theta)$, is used to calculate the penalty function.

Calculating and minimizing the final penalty function

First the penalty function (see equation 3.1) is calculated for each realization and each set of tolerance parameters. After calculating these penalties the final penalty, $\bar{p}(\theta)$ which is function of tolerance parameters is calculated by averaging the calculated penalties over different realizations for the fixed tolerance parameters. Therefore

$$\bar{p}(\theta) = \frac{1}{n_{realz}} \cdot \sum_{i_{realz}=1}^{n_{realz}} p_{i_{realz}}(\theta) \quad \text{.....(3.36)}$$

where n_{realz} is the number of realizations, θ corresponds to the fixed tolerance parameters and $p_{i_{realz}}(\theta)$ is the penalty for realization i_{realz} and is calculated by using equation 3.1.

After calculating the final penalty, it should be minimized to get the optimal tolerance parameters. The tolerance parameters have different number of components for different cases. For example for omnidirectional 2D and 3D data sets, there are two components for variogram calculation, they are unit lag distance and the lag tolerance (instead of lag tolerance the tolerance ratio can also be used). For other general cases the number of tolerance components is greater than 2.

For minimizing the penalty function in the presence of multiple variables a sequential type approach could be used to obtain the optimal point which minimizes the penalty function. For example if there are three variables, at first step two of them should be fixed and the penalty should be minimized with respect the variable which is not fixed. After specifying this point the next variable should be optimized and the others should be fixed. This procedure can be applied step by step to get the final optimal point and as mentioned

at each step one of the variables should be optimized and the others are fixed and each of the variables is optimized once.

After obtaining the optimal tolerance parameters, the experimental variogram should be calculated by using the optimal tolerance parameters and fitted with the `varfit` program.

3.2.2 Calculating the tolerance correction factor

Calculating the experimental variogram by using some tolerance parameters affects the estimated value of the maximum and minimum range of continuity (after fitting). It can be proved mathematically that for example in 2D case, the true maximum range of continuity is underestimated and the true minimum range of continuity is overestimated. The amount of increase/decrease refers to the tolerance correction factor. So in general case the calculated experimental points are artificially shifted to the left in major direction and to the right in minor direction. The analytical formulas for 2D and 3D cases are presented in this section; the proofs for the analytical relationship are given in Appendix A.

3.2.2.1 Tolerance correction factor in 2D without bandwidth

Assume that the ellipse in Figure 3.6 shows the anisotropy of the variable in a specific 2D field. The field has a maximum range of correlation of $a_{h-\max}$ (the maximum radius of the ellipse) and the minimum range of correlation of $a_{h-\min}$. If we introduce an angle tolerance of a_{tol}, then the tolerance will cause an underestimation of $a_{h-\max}$ and an overestimation of $a_{h-\min}$. Therefore the anisotropy ratio will artificially decrease just because of introducing tolerance parameters. For example if $a_{h-\max}=4$ and $a_{h-\min}=1$ (the anisotropy ratio would be 4:1) then introducing an angle tolerance of 22.5 degree will underestimate $a_{h-\max}$ to 3.18 and overestimate $a_{h-\min}$ to 1.02 and the apparent anisotropy ratio would be 3.1:1.

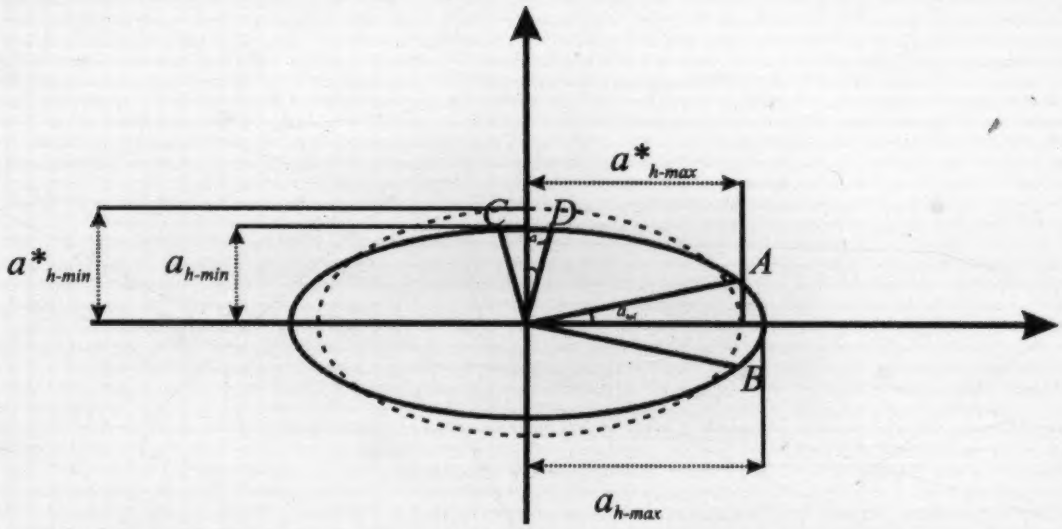


Figure 3.6. True Anisotropic ellipse (solid) and the estimated anisotropic ellipse (dashed) in 2D case when there is no bandwidth

The relationship between the true range of correlations ($a_{h-\min}$, $a_{h-\max}$) and the estimated range of correlations ($a^*_{h-\min}$, $a^*_{h-\max}$) is derived. The estimated value for the range of correlation is the average of the radii in the tolerance region over the tolerance region. Therefore the estimated values can be written as:

$$a^*_{h-\max} = \frac{\int_{-a_{tol}}^{a_{tol}} r d\theta}{\int_{-a_{tol}}^{a_{tol}} d\theta} \quad \text{.....(3.37)}$$

$$a^*_{h-\min} = \frac{\int_{\frac{\pi}{2}-a_{tol}}^{\frac{\pi}{2}+a_{tol}} r d\theta}{\int_{\frac{\pi}{2}-a_{tol}}^{\frac{\pi}{2}+a_{tol}} d\theta} \quad \text{.....(3.38)}$$

By using the equation for the ellipse in polar coordinate the integrals can be solved, therefore:

$$a^{*}_{h-\max} = a_{h-\min} \cdot \frac{F\left(\frac{\pi}{2},\omega\right) - F\left(\frac{\pi}{2} - a_{tol},\omega\right)}{a_{tol}} < a_{h-\max} \quad \text{.....(3.39)}$$

$$a^{*}_{h-\min} = a_{h-\min} \cdot \frac{F\left(a_{tol},\omega\right)}{a_{tol}} > a_{h-\min} \quad \text{.....(3.40)}$$

Where ω is the eccentricity of the ellipse and has a value between 0 and 1 and calculated as below:

$$\omega = \sqrt{1 - \frac{a^2_{h-\min}}{a^2_{h-\max}}} \quad ; \quad 0 < a_{h-\min} \le a_{h-\max} \quad \text{.....(3.41)}$$

And $F(\alpha,\omega)$ is the incomplete Legendre elliptic integral of the 1st kind and is defined as below (Abramowitz and Stegun, 1965):

$$F(\alpha,\omega) = \int_0^{\alpha} \frac{d\theta}{\sqrt{1-\omega^2 \sin^2\theta}} \quad \text{.....(3.42)}$$

The values for the incomplete Legendre elliptic integral of the 1st kind are tabulated in many mathematical handbooks (e.g. Abramowitz and Stegun, 1965). This function can be also calculated numerically by using the code provided in Numerical Recipes in Fortran 77 (Press *et al*, 1992).

By using the obtained formulas for $a^{*}_{h-\min}$ and $a^{*}_{h-\max}$, the estimated anisotropy ratio can be written as below:

$$\frac{a^{*}_{h-\max}}{a^{*}_{h-\min}} = \frac{F\left(\frac{\pi}{2},\omega\right) - F\left(\frac{\pi}{2} - a_{tol},\omega\right)}{F\left(a_{tol},\omega\right)} \quad \text{.....(3.43)}$$

We know that the true anisotropy ratio is $\frac{a_{h-\max}}{a_{h-\min}} = \frac{1}{\sqrt{1-\omega^2}}$, since $a_{h-\min}$ is overestimated and $a_{h-\max}$ is underestimated therefore $\frac{a^*_{h-\max}}{a^*_{h-\min}}$ will be decreased, therefore:

$$\frac{a_{h-\max}}{a_{h-\min}} > \frac{a^*_{h-\max}}{a^*_{h-\min}} \quad \text{.....(3.44)}$$

It can be proved that, $F\left(\frac{\pi}{2},\omega\right) - F\left(\frac{\pi}{2} - a_{tol},\omega\right) = F\left(\arctan\left(\frac{\tan a_{tol}}{\sqrt{1-\omega^2}}\right),\omega\right)$ (Abramowitz and Stegun, 1965) therefore the equations 3.39 and 3.40 can be written as below:

$$a^*_{h-\max} = a_{h-\min} \cdot \frac{F\left(\arctan\left(\frac{\tan a_{tol}}{\sqrt{1-\omega^2}}\right),\omega\right)}{a_{tol}} \quad \text{.....(3.45)}$$

$$\frac{a^*_{h-\max}}{a^*_{h-\min}} = \frac{F\left(\arctan\left(\frac{\tan a_{tol}}{\sqrt{1-\omega^2}}\right),\omega\right)}{F\left(a_{tol},\omega\right)} \quad \text{.....(3.46)}$$

For example if $a_{h-\max} = 4$, $a_{h-\min} = 1$, $a_{tol} = 22.5^\circ = \frac{\pi}{8}$ then the eccentricity is, $\omega = \frac{\sqrt{15}}{4}$ and by using the derived formulas for the estimated ranges of correlations, we have:

$$a^*_{h-\max} = 1 \times \frac{F\left(\frac{\pi}{2},\frac{\sqrt{15}}{4}\right) - F\left(\frac{3\pi}{8},\frac{\sqrt{15}}{4}\right)}{\left(\frac{\pi}{8}\right)} = 3.184601$$

$$a^*_{h-\min} = 1 \times \frac{F\left(\frac{\pi}{8},\frac{\sqrt{15}}{4}\right)}{\left(\frac{\pi}{8}\right)} = 1.024952$$

It can be seen that $a^*_{h-\min} > a_{h-\min}$ and $a^*_{h-\max} < a_{h-\max}$ therefore the anisotropy ratio is also decreased:

$$\frac{a^{*}_{h-\max}}{a^{*}_{h-\min}} = \frac{3.184601}{1.024952} \cong 3.107073 < \frac{a_{h-\max}}{a_{h-\min}} = 4 \quad \Rightarrow \quad \frac{a^{*}_{h-\max}}{a^{*}_{h-\min}} < \frac{a_{h-\max}}{a_{h-\min}}$$

The more important problem in the case of the variogram calculations is what the true values for the minor and major direction of continuity are if we know the estimated values for them. In almost all of the cases we know the estimated values. To do that we should solve the derived relations for calculating the $a_{h-\min}$ and $a_{h-\max}$ by knowing, $a^{*}_{h-\min}$, $a^{*}_{h-\max}$ and a_{tol}.

We know that:

$$\frac{a^{*}_{h-\max}}{a^{*}_{h-\min}} = \frac{F\left(\arctan\left(\frac{\tan a_{tol}}{\sqrt{1-\omega^2}}\right), \omega\right)}{F\left(a_{tol}, \omega\right)}$$

After rearranging:

$$f(\omega) = \left(\frac{a^{*}_{h-\max}}{a^{*}_{h-\min}}\right) \cdot F\left(a_{tol}, \omega\right) - F\left(\arctan\left(\frac{\tan a_{tol}}{\sqrt{1-\omega^2}}\right), \omega\right) = 0 \quad \text{.....(3.47)}$$

The values for, $a^{*}_{h-\min}$, $a^{*}_{h-\max}$ and a_{tol} are known therefore the above equation can be solved to calculate ω which has the value between 0 and 1. Bisection method can be used for finding the root of $f(\omega)$. After finding the actual eccentricity, the true $a_{h-\min}$ and $a_{h-\max}$ can be calculated as below:

$$a_{h-\min} = \frac{a^{*}_{h-\min} \cdot a_{tol}}{F\left(a_{tol}, \omega\right)} \quad \text{.....(3.48)}$$

$$a_{h-\max} = \frac{a_{h-\min}}{\sqrt{1-\omega^2}} \quad \text{.....(3.49)}$$

3.2.2.2 *Tolerance correction factor in 2D with bandwidth*

In this case since a bandwidth is introduced therefore two different apparent angle tolerances should be used instead of the true angle tolerance in both of the minimum and maximum directions of continuity. These two apparent angle tolerances are calculated as below. The apparent angle tolerance for maximum direction, $a^{app}_{tol,\max}$, is shown in Figure 3.7.

$$a^{app}_{tol,\max} = \arctan\left(\frac{b}{a_{h-\max}\sqrt{1-\frac{b^2}{a^2_{h-\min}}}}\right) \quad \text{.....(3.50)}$$

$$a^{app}_{tol,\min} = \arctan\left(\frac{b}{a_{h-\min}\sqrt{1-\frac{b^2}{a^2_{h-\max}}}}\right) \quad \text{.....(3.51)}$$

where b is the bandwidth.

It is obvious that b should have a maximum value, if we want to see its effect on variogram calculation in the presence of anisotropic ellipse. For calculating $a^{app}_{tol,\max}$, the maximum value of b is equal to $b^*_{\min}$. $b^*_{\min}$ is indicated in Figure 3.7. It is function of the a_{tol}, $a_{h-\min}$ and $a_{h-\max}$. $b^*_{\min}$ is calculated as:

$$b^*_{\min} = \frac{\tan a_{tol}}{\sqrt{\frac{1}{a^2_{h-\max}} + \frac{\tan^2 a_{tol}}{a^2_{h-\min}}}} \quad \text{.....(3.52)}$$

The bandwidth, b, in this case is satisfied in below inequality:

$$b \le b^*_{\min} \le a_{h-\min}$$

And for calculating $a^{app}_{tol,\min}$, the maximum value of b is equal to $b^*_{\max}$. It is function of the a_{tol}, $a_{h-\min}$ and $a_{h-\max}$. $b^*_{\max}$ is calculated as:

$$b^{*}_{\min} = \frac{\tan a_{tol}}{\sqrt{\frac{\tan^2 a_{tol}}{a^2_{h-\max}} + \frac{1}{a^2_{h-\min}}}} \quad \text{.....(3.53)}$$

The bandwidth, b, in this case is satisfied in below inequality:

$$b \le b^{*}_{\max} \le a_{h-\max}$$

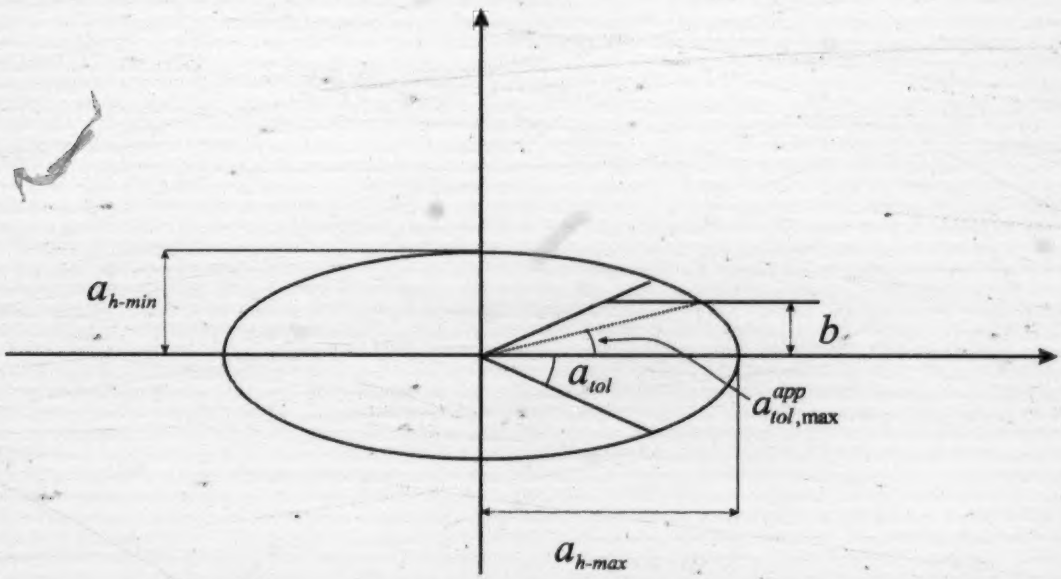


Figure 3.7. Anisotropic ellipse in 2D case with bandwidth for maximum direction of continuity

The formulas for estimated range of correlation in two major directions in the case of bandwidth reduce to:

$$a^{*}_{h-\max} = a_{h-\min} \cdot \frac{F\left(\frac{\pi}{2}, \omega\right) - F\left(\frac{\pi}{2} - a^{app}_{tol,\max}, \omega\right)}{a^{app}_{tol,\max}} \quad \text{.....(3.54)}$$

$$a^{*}_{h-\min} = a_{h-\min} \cdot \frac{F\left(a^{app}_{tol,\min}, \omega\right)}{a^{app}_{tol,\min}} > a_{h-\min} \quad \text{.....(3.55)}$$

To obtain $a_{h-\min}$ and $a_{h-\max}$ a system of non-linear equations should be solved. In addition to above two formulas for $a_{h-\min}$ and $a_{h-\max}$, there are three auxiliary equations for

calculating $a_{tol,\min}^{app}$, $a_{tol,\max}^{app}$ and ω which were given before. The known parameters are the values for the bandwidth, angle tolerance and the estimated ranges.

3.2.2.3 Tolerance correction factor in 3D without bandwidth

In 3D tolerance case in addition to obtaining the relation between the true and the estimated ranges in horizontal (maximum and minimum directions of continuity) direction, the relationship should be obtained for vertical direction as well. The idea for calculating the estimated ranges are the same as in 2D case but the relations are more complicated. The required integrals are written in spherical coordinates, therefore for horizontal direction we have:

$$a_{h-\max}^{*} = \frac{\int_{-a_{tol}^{h}}^{a_{tol}^{h}} \int_{\phi_1(\theta)}^{\pi-\phi_1(\theta)} \rho d\phi d\theta}{\int_{-a_{tol}^{h}}^{a_{tol}^{h}} \int_{\phi_1(\theta)}^{\pi-\phi_1(\theta)} d\phi d\theta} \quad \text{.....(3.56)}$$

$$a_{h-\min}^{*} = \frac{\int_{\frac{\pi}{2}-a_{tol}^{h}}^{\frac{\pi}{2}+a_{tol}^{h}} \int_{\phi_2(\theta)}^{\pi-\phi_2(\theta)} \rho d\phi d\theta}{\int_{\frac{\pi}{2}-a_{tol}^{h}}^{\frac{\pi}{2}+a_{tol}^{h}} \int_{\phi_2(\theta)}^{\pi-\phi_2(\theta)} d\phi d\theta} \quad \text{.....(3.57)}$$

And for vertical direction:

$$a_{ver}^{*} = \frac{\int_{0}^{2\pi} \int_{0}^{a_{tol}} \rho d\phi d\theta}{\int_{0}^{2\pi} \int_{0}^{a_{tol}} d\phi d\theta} \quad \text{.....(3.58)}$$

Where ρ is the distance from the centre of the ellipsoid to its circumference in spherical coordinate and $\phi_1(\theta)$ and $\phi_2(\theta)$ are:

$$\phi_1(\theta) = \operatorname{arccot}\left(\tan^2 a_{tol}^v \cdot \sqrt{1-\left(1+\cot^2 a_{tol}^h\right)\cdot \sin^2\theta}\right) \quad \text{.....(3.59)}$$

$$\phi_2(\theta) = \operatorname{arccot}\left(\tan^2 a_{tol}^v \cdot \sqrt{1-\left(1+\cot^2 a_{tol}^h\right)\cdot \cos^2\theta}\right) \quad \text{.....(3.60)}$$

After integrating and applying the boundaries of the integrals for horizontal direction we will have:

$$a_{h-\max}^{*} = \frac{a_{ver}}{\pi \cdot a_{tol}^h - 2\int_0^{a_{tol}^h} \phi_1(\theta)\cdot d\theta} \cdot \int_0^{a_{tol}^h} \left[F\left(\pi-\phi_1(\theta),\omega_1(\theta)\right) - F\left(\phi_1(\theta),\omega_1(\theta)\right)\right]\cdot d\theta \quad \text{.....(3.61)}$$

$$a_{h-\min}^{*} = \frac{a_{ver}}{\pi \cdot a_{tol}^h - 2\int_0^{a_{tol}^h} \phi_1(\theta)\cdot d\theta} \cdot \int_0^{a_{tol}^h} \left[F\left(\pi-\phi_1(\theta),\omega_2(\theta)\right) - F\left(\phi_1(\theta),\omega_2(\theta)\right)\right]\cdot d\theta \quad \text{.....(3.62)}$$

And for vertical direction we have:

$$a_{ver}^{*} = \frac{a_{ver}}{2\pi \cdot a_{tol}} \cdot \int_0^{2\pi} F\left(a_{tol},\omega_1(\theta)\right)\cdot d\theta \quad \text{.....(3.63)}$$

Where $\omega_1(\theta)$ and $\omega_2(\theta)$ are calculated as:

$$\omega_1(\theta) = 1-\left(\frac{a_{ver}^2}{a_{h-\max}^2}\right)-\left(\frac{a_{ver}^2}{a_{h-\min}^2 \cdot a_{h-\max}^2}\right)\cdot\left(a_{h-\max}^2 - a_{h-\min}^2\right)\cdot \sin^2\theta \quad \text{.....(3.64)}$$

$$\omega_2(\theta) = 1-\left(\frac{a_{ver}^2}{a_{h-\max}^2}\right)-\left(\frac{a_{ver}^2}{a_{h-\min}^2 \cdot a_{h-\max}^2}\right)\cdot\left(a_{h-\max}^2 - a_{h-\min}^2\right)\cdot \cos^2\theta \quad \text{.....(3.65)}$$

In the case that the angle tolerances and the estimated values for the ranges are known then by solving the three non-linear equations (equations for $a_{h-\min}^{*}$, $a_{h-\max}^{*}$ and a_{ver}^{*}) simultaneously, the three unknowns($a_{h-\min}$, $a_{h-\max}$ and a_{ver}) can be determined.

3.2.2.4 *Tolerance correction factor in 3D with bandwidth*

In the case of bandwidth in 3D, for horizontal direction, two cases might happen:

1. There is vertical bandwidth (b_{ver}) but no horizontal bandwidth (b_{hor}) is applied.

   In this case the vertical bandwidth introduces two apparent angle tolerances, $a_{tol,\min}^{v,app}$ and $a_{tol,\max}^{v,app}$ instead of a_{tol}^{v}. These two apparent angles are different because of the minimum and maximum directions of continuity and calculated as:

$$a_{tol,\min}^{v,app} = \arctan\left(\frac{b_{ver}}{a_{h-\min}\sqrt{1-\frac{b_{ver}^2}{a_{ver}^2}}}\right) \quad \text{.....(3.66)}$$

$$a_{tol,\max}^{v,app} = \arctan\left(\frac{b_{ver}}{a_{h-\max}\sqrt{1-\frac{b_{ver}^2}{a_{ver}^2}}}\right) \quad \text{.....(3.67)}$$

   The equations for $a_{h-\min}^{*}$ and $a_{h-\max}^{*}$ are

$$a_{h-\max}^{*} = \frac{a_{ver}}{\pi \cdot a_{tol}^{h} - 2\int_{0}^{a_{tol}^{h}} \phi_1(\theta)\cdot d\theta} \cdot \int_{0}^{a_{tol}^{h}} \left[F\left(\pi-\phi_1(\theta),\omega_1(\theta)\right)-F\left(\phi_1(\theta),\omega_1(\theta)\right)\right]\cdot d\theta \ldots$$

.....(3.68)

$$a_{h-\min}^{*} = \frac{a_{ver}}{\pi \cdot a_{tol}^{h} - 2\int_{0}^{a_{tol}^{h}} \phi_2(\theta)\cdot d\theta} \cdot \int_{0}^{a_{tol}^{h}} \left[F\left(\pi-\phi_2(\theta),\omega_2(\theta)\right)-F\left(\phi_2(\theta),\omega_2(\theta)\right)\right]\cdot d\theta$$

.....(3.69)

   where $\phi_1(\theta)$ and $\phi_2(\theta)$ are calculated as

$$\phi_1(\theta) = \operatorname{arccot}\left(\tan^2 a_{tol,\max}^{v,app}\cdot\sqrt{1-\left(1+\cot^2 a_{tol}^{h}\right)\cdot\sin^2\theta}\right) \quad \text{.....(3.70)}$$

$$\phi_2(\theta) = \operatorname{arccot}\left(\tan^2 a_{tol,\min}^{v,app} \cdot \sqrt{1-\left(1+\cot^2 a_{tol}^{h}\right)\cdot\cos^2\theta}\right) \quad \text{.....(3.71)}$$

2. Both of the horizontal and vertical bandwidths are present.

In this case four apparent angle tolerances are introduced two in minimum directions, $a_{tol,\min}^{v,app}$ and $a_{tol,\min}^{h,app}$, and two in maximum directions, $a_{tol,\max}^{v,app}$ and $a_{tol,\max}^{h,app}$. The equations for $a_{tol,\min}^{v,app}$ and $a_{tol,\max}^{v,app}$ are the same as before but for $a_{tol,\min}^{h,app}$ and $a_{tol,\max}^{h,app}$ we have:

$$a_{tol,\min}^{h,app} = \arctan\left(\frac{b_{hor}}{a_{h-\min}\sqrt{1-\dfrac{b_{hor}^2}{a_{h-\max}^2}}}\right) \quad \text{.....(3.72)}$$

$$a_{tol,\max}^{h,app} = \arctan\left(\frac{b_{hor}}{a_{h-\max}\sqrt{1-\dfrac{b_{hor}^2}{a_{h-\min}^2}}}\right) \quad \text{.....(3.73)}$$

And finally the formulas for $a_{h-\min}^{\bullet}$ and $a_{h-\max}^{\bullet}$ are

$$a_{h-\max}^{\bullet} = \frac{a_{ver}}{\pi\cdot a_{tol,\max}^{h,app} - 2\int_0^{a_{tol,\max}^{h,app}} \phi_1(\theta)\cdot d\theta} \cdot \int_0^{a_{tol,\max}^{h,app}} \left[F\left(\pi-\phi_1(\theta),\omega_1(\theta)\right) - F\left(\phi_1(\theta),\omega_1(\theta)\right)\right]\cdot d\theta$$

.....(3.74)

$$a_{h-\min}^{\bullet} = \frac{a_{ver}}{\pi\cdot a_{tol,\min}^{h,app} - 2\int_0^{a_{tol,\min}^{h,app}} \phi_2(\theta)\cdot d\theta} \cdot \int_0^{a_{tol,\min}^{h,app}} \left[F\left(\pi-\phi_2(\theta),\omega_2(\theta)\right) - F\left(\phi_2(\theta),\omega_2(\theta)\right)\right]\cdot d\theta$$

.....(3.75)

where $\phi_1(\theta)$ and $\phi_2(\theta)$ are calculated as

$$\phi_1(\theta) = \operatorname{arccot}\left(\tan^2 a_{tol,\max}^{v,app} \cdot \sqrt{1-\left(1+\cot^2 a_{tol,\max}^{h,app}\right)\cdot\sin^2\theta}\right) \quad \text{.....(3.76)}$$

$$\phi_2(\theta) = \operatorname{arccot}\left(\tan^2 a_{tol,\min}^{v,app} \cdot \sqrt{1-\left(1+\cot^2 a_{tol,\min}^{h,app}\right)\cdot\cos^2\theta}\right) \quad (3.77)$$

3.2.3 Rescaling the experimental variogram points

In 2D case, the tolerance correction factor in the two minor and major directions of continuity can be calculated as:

$$f_{\min} = \frac{a_{h-\min}}{a^{*}_{h-\min}} \quad (3.78)$$

$$f_{\max} = \frac{a_{h-\max}}{a^{*}_{h-\max}} \quad (3.79)$$

These factors in minor and major directions should be applied on all of the calculated experimental points therefore the lag distance axis on variogram plot for major and minor directions should be rescaled:

$$\mathbf{h}_{new} = f_{\min} \cdot \mathbf{h} \quad (3.80)$$

$$\mathbf{h}_{new} = f_{\max} \cdot \mathbf{h} \quad (3.81)$$

It should be noted that the variogram values are not changed, the plot is just shifted horizontally to the left or to the right depending on the minor or major directions of continuity.

For 3D the optimal experimental lag distances should be updated and a variogram should be fitted again. The new lag distance should be obtained for three different directions: minor, major and vertical directions. The tolerance correction factors for minor and major directions are the same as equations 3.78 and 3.79, respectively. For the vertical direction, the correction factor is:

$$f_{ver} = \frac{a_{ver}}{a^{*}_{ver}} \quad (3.82)$$

And the corresponding rescaled lag distance in vertical direction is:

$$\mathbf{h}_{new} = f_{ver} \cdot \mathbf{h} \quad (3.83)$$

After rescaling the lag distance axis in all of three directions and updating the experimental points, a variogram should be fitted by using `varfit`.

Chapter 4

Application

4.1 Synthetic case

In this synthetic example a 2D data set is considered and we are trying to find the optimum parameters for calculation of experimental omnidirectional variogram, the parameters which can be changed are the lag distance, h and lag tolerance, it should be noted that the angle tolerance is 90 degree and the bandwidth is set to large number to get the omnidirectional variogram.

Problem Setting

A synthetic 2D Gaussian random field (Figure 4.1) is generated via an unconditional simulation for a 1024 x 1024 grid with the following isotropic variogram as a reference variogram:

$$\gamma(\mathbf{h}) = 0.05 + 0.95 Sph_{a=64}(\|\mathbf{h}\|)$$

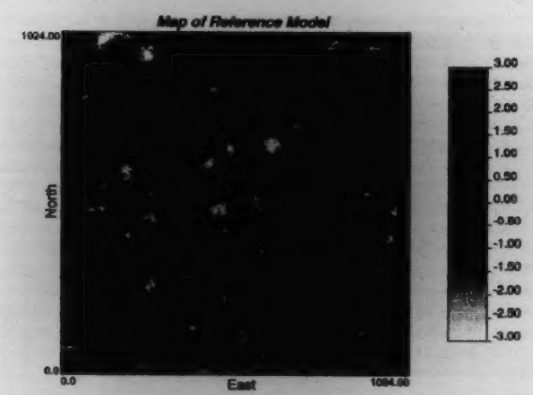


Figure 4.1. Map of reference model generated by `sgsim`

$Sph_{a=64}(\|\mathbf{h}\|)$ is the isotropic spherical structure with a range of 64. The mathematical definition of $Sph_{a=64}(\|\mathbf{h}\|)$ is:

$$Sph_{a=64}(\|\mathbf{h}\|) = \begin{cases} 1.5\times\left(\frac{\|\mathbf{h}\|}{64}\right) - 0.5\times\left(\frac{\|\mathbf{h}\|}{64}\right)^3 & , \text{if } \|\mathbf{h}\| < 64 \\ 1 & , \text{if } \|\mathbf{h}\| \geq 64 \end{cases}$$

$\|\mathbf{h}\|$ is the norm of vector $\mathbf{h}$.

To make sure that the result has a right variogram that we have used to build it, variogram reproduction is checked. Figure 4.2 shows the variogram reproduction.

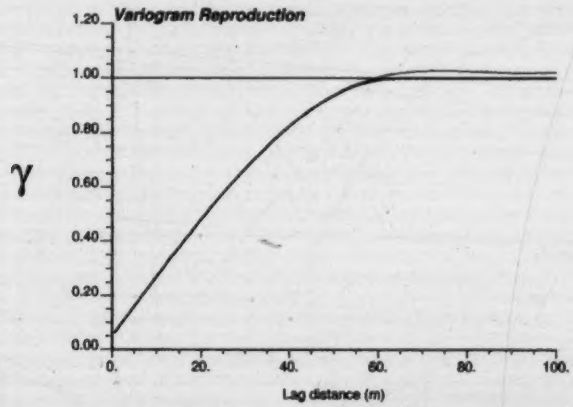


Figure 4.2. Variogram reproduction the reference model generated by `sgsim`

Using this reference model, we can then sample randomly at n locations (n=200 for this example) by using the `draw` program (Deutsch and Journel, 1998). The locations of the 200 picked points are shown in Figure 4.3.

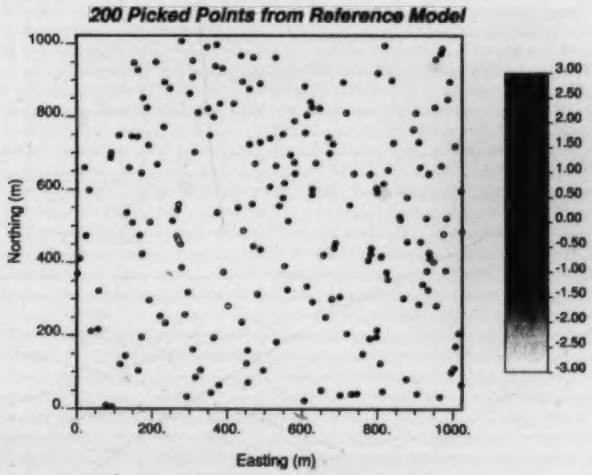


Figure 4.3. Locations of 200 randomly picked points from reference model

Figure 4.4 shows the reference distribution which is the Gaussian distribution and also the distribution of the 200 picked points.

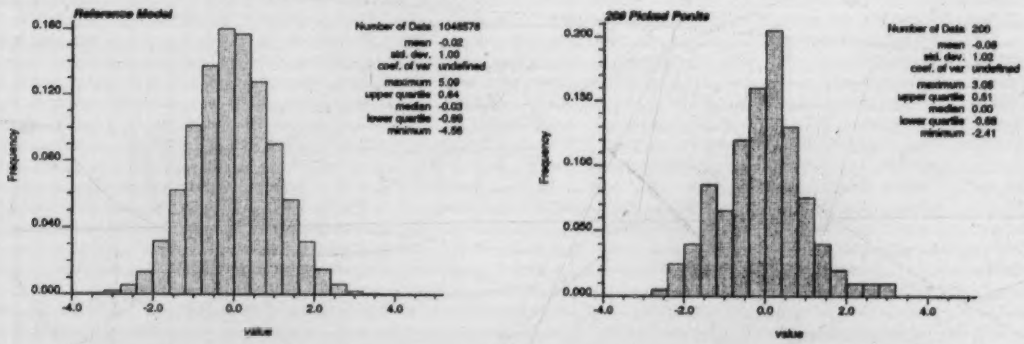


Figure 4.4. Reference Gaussian distribution (left) and the distribution of the 200 randomly picked points (right), the values are in Gaussian units

Now the data set is constructed, the only difference of this data set with the real data set is that we know the reference variogram model of the field for this case. Spatial bootstrap can be used to assess uncertainty associated with variogram calculation. It gives N realization at each location, by using N sets of data; N different experimental variograms can be calculated, so at each lag there are N values for variogram which can be used to make the variogram distribution at each lag, therefore the variogram uncertainty can be quantified. Figure 4.5 shows the reference variogram and also the experimental points

based on the 200 randomly picked points. Our goal is to quantify the uncertainty in the variogram at the lags that the experimental variogram is calculated.

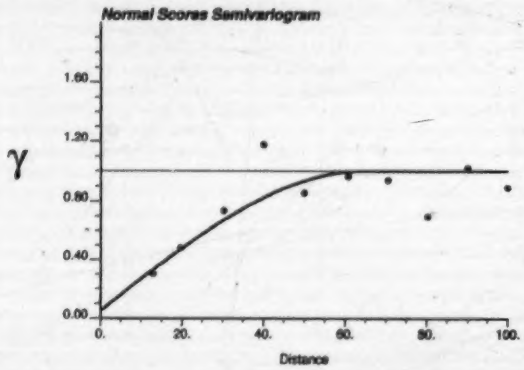


Figure 4.5. The Reference variogram (solid line) and the experimental variogram points based on 200 picked points (bullets), the distance is in meters (m)

By using spatial bootstrap method, different realizations (e.g. 100 realizations) can be calculated at each of the 200 data locations. For each realization (which consists of 200 simulated values), the experimental variogram can be calculated. Therefore at each lag distance we have 100 values (realizations) for the variogram. Figure 4.6 shows the distribution of the experimental variogram at each lag distance and the distribution of the variogram values at **h**=40. The distribution of the experimental variogram values are plotted at each lag distance in Figure 4.7. At each lag the variance of the variogram is calculated and it is plotted against the lag distance **h**. Figure 4.8 shows the plot of the variance of the variogram from spatial bootstrap against lag distance.

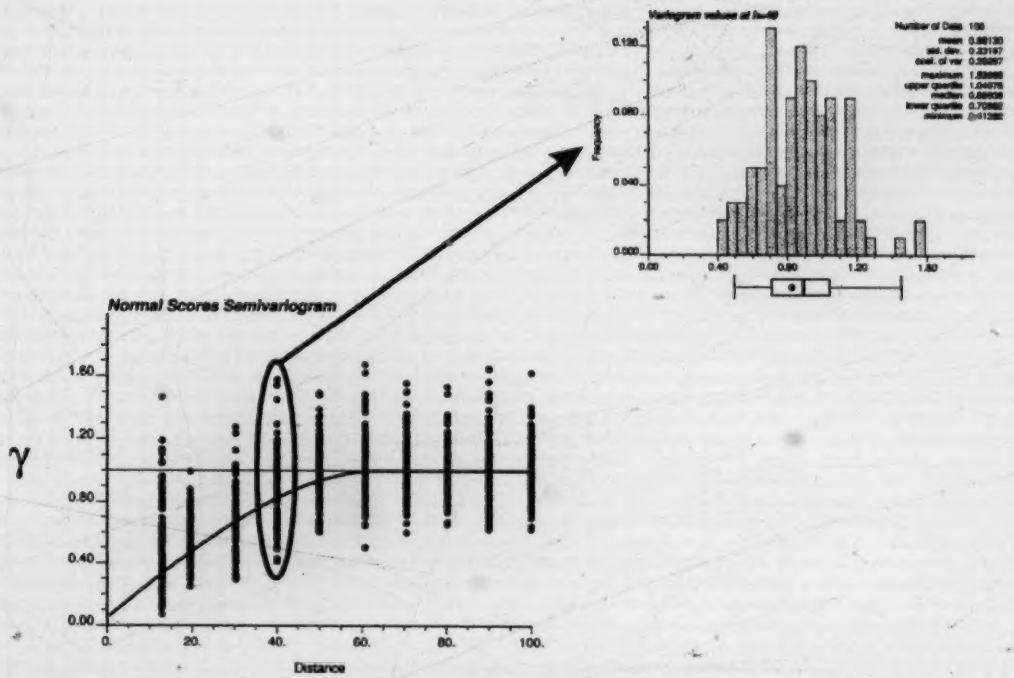


Figure 4.6. Different realizations for variogram at each lag accompanying by the probability density function at **h**=40

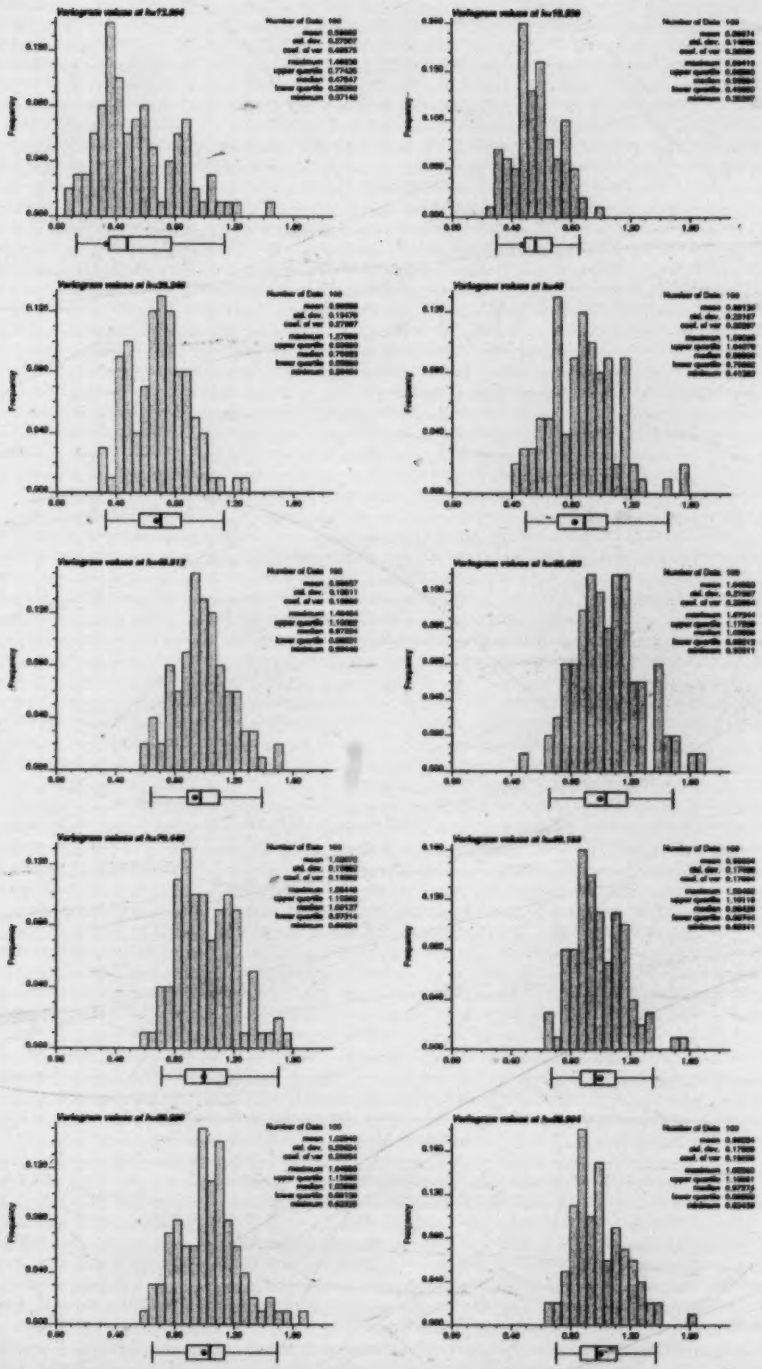


Figure 4.7. Probability density function of the variogram values at each lag distances (see Figure 4.6), for each histogram a box plot is shown, the black dot is the reference variogram value

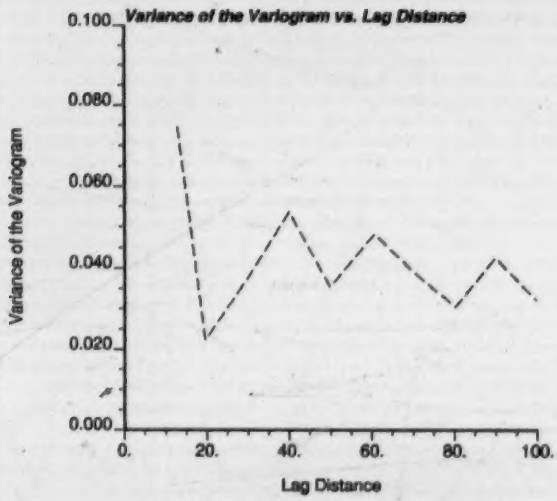


Figure 4.8. Variance of the variogram versus the lag distance (m)

Below procedures are performed to find the optimum parameters for calculation of omnidirectional variogram:

1. Using spatial bootstrap to obtain L values (realizations) for each location; e.g. L=100.

2. Calculating experimental variogram for each realization with different tolerance parameters (lag separation distance, tolerance ratio and the number of lags).

$$h = 2,4,6,...,50$$

$$Tol\ \% = \frac{h_{tol}}{h} \times 100 = 4,8,12,...,100$$

$$n_{lag} = \text{int}\left(\frac{|A|}{2h}\right) = \text{int}\left(\frac{512}{h}\right)$$

3. For each calculated experimental variogram a variogram should be fitted by using `varfit` program. Fitting variogram will reduce the uncertainty in the variogram. This can be shown by using spatial bootstrap before and after variogram fitting. At each lag distance there are 100 realizations as a result of spatial bootstrap therefore there are 100 variogram fits for them. Figure 4.9 shows the variogram plot before and after variogram fitting, it can be seen that the range of variation

for the variogram values is reduced after fitting. Figure 4.10 shows the reduction in the variance of the variogram as a result of variogram fitting.

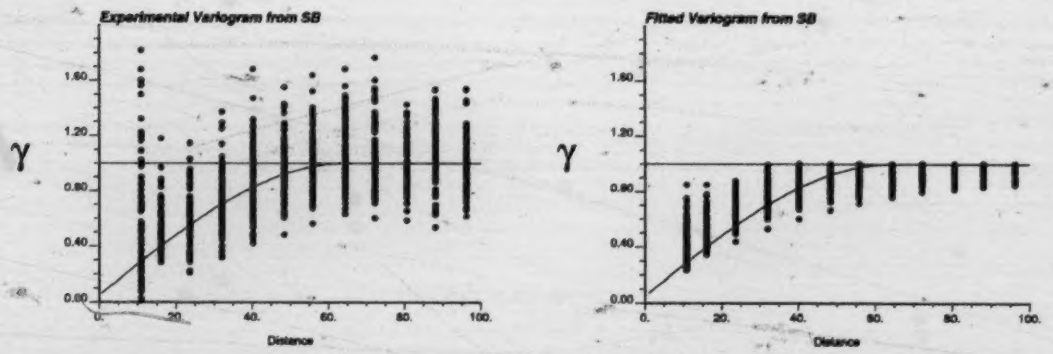


Figure 4.9. Calculated experimental variogram for 100 realizations from Spatial Bootstrap at h=8 and h_{tol}=4.48 (left) and the corresponding fitted variogram values (right)

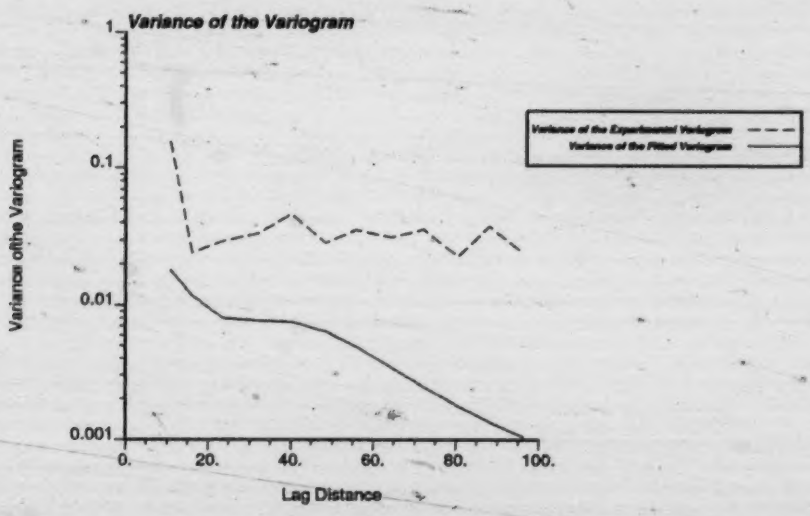


Figure 4.10. Variance of the experimental variogram (dash line) and the variance of the fitted variogram (solid line) versus the lag distance

4. By using the fitted variogram in step 3 and the reference variogram a penalty function, P, can be calculated by using equation 3.1.

5. Plotting the final penalty versus the lag separation distance and the tolerance ratio. The final penalty is the average of the penalties which are calculated in previous step over different realizations. Figure 4.11 shows the final penalty versus the tolerance parameters. The conditional expectation of the final penalty is plotted in Figure 4.12.

6. Trying to find the optimum lag separation distance and the lag tolerance which minimize the final penalty. The optimum values happen at h=18 and Tol=56 %.

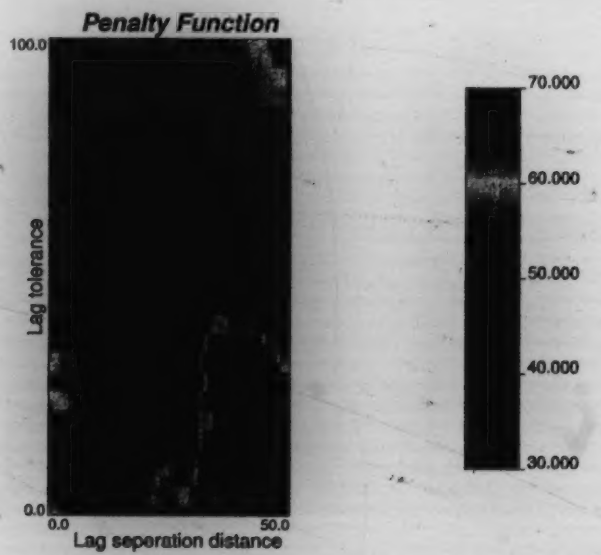


Figure 4.11. Final penalty as a function of both lag tolerance (m) and lag separation distance (m)

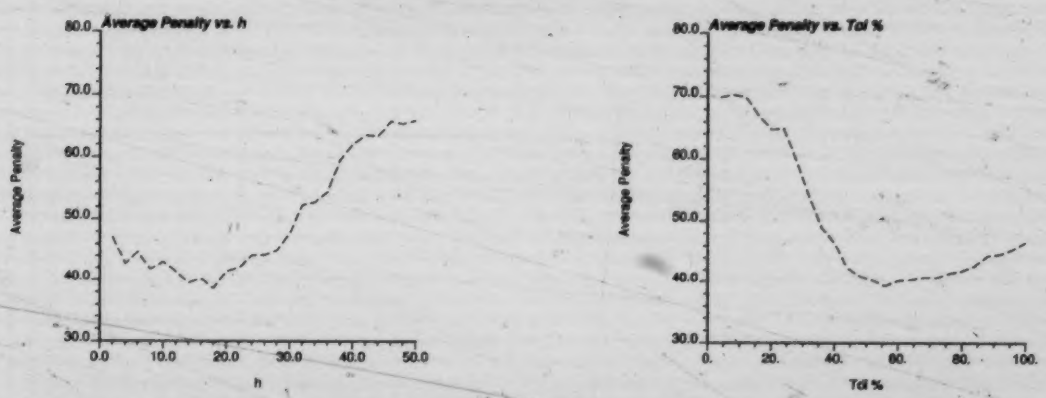


Figure 4.12. Conditional expectation of the average penalty versus the unit lag distance (left) and the tolerance ratio (right)

Effect of number of randomly picked points and field size on optimal tolerance parameters

The synthetic case that was presented already gives us the optimal tolerance parameters for a certain example. But it is obvious that the optimal tolerance parameters should be function of the field size, number of data and the reference variogram parameters (nugget effect, range and the type of variogram model, i.e. spherical). To see the effect of number of data and field size on optimal tolerance parameters, different cases were run with field size of (512x512 and 1024x1024) and reference variogram parameters (isotropic 2D spherical variogram with the nugget effect of 0.05 and the range of 64) but with different number of randomly picked data points (n=128, 256, 512 and 1024). The reference 2D map for 1024x1024 is the same as Figure 4.2 but for 512x512 another 2D Gaussian field is generated by using `sgsim`. The reference map for this case is shown in Figure 4.13. The variogram reproduction is checked for this case and is plotted in Figure 4.14. The locations of the different randomly picked points are shown in Figure 4.15. To see better the effect of number of data, for each specific number of randomly picked data points and field size different realizations are considered. The realizations are considered by changing the random seed number in `draw` program. Figure 4.16 shows four different realizations for 256 randomly picked data points for a fixed field size of 1024.

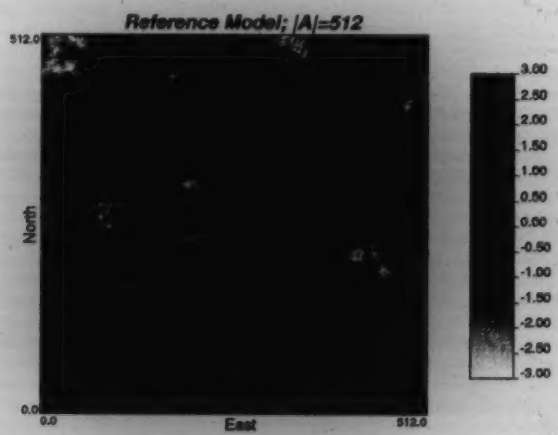


Figure 4.13. Map of reference model generated by `sgsim` for field size of 512

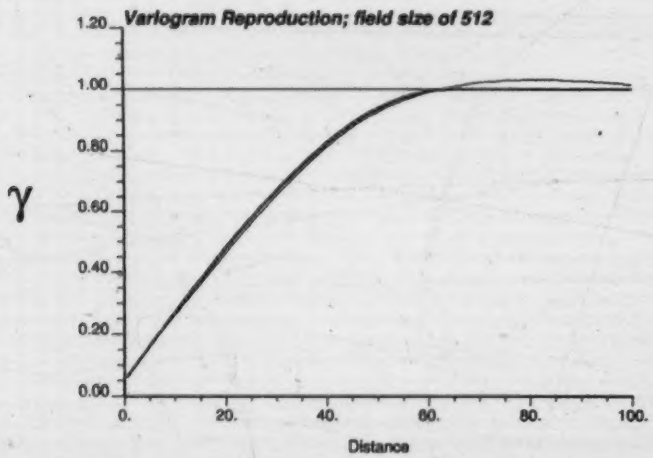


Figure 4.14. Variogram reproduction check for the reference model with field size of 512, distance is in meters (m)

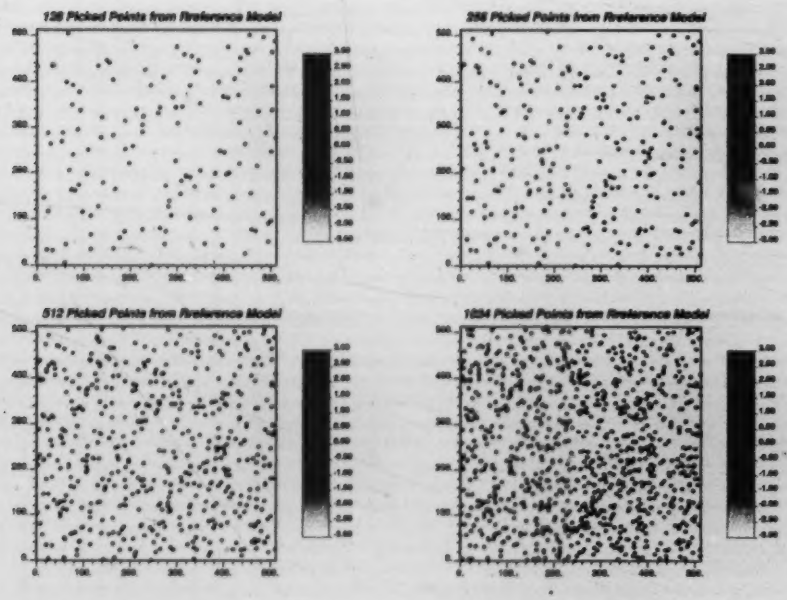

128 Picked Points from Reference Model
256 Picked Points from Reference Model
512 Picked Points from Reference Model
1024 Picked Points from Reference Model

Figure 4.15. Locations of different number of randomly picked points from reference model with field size of 512

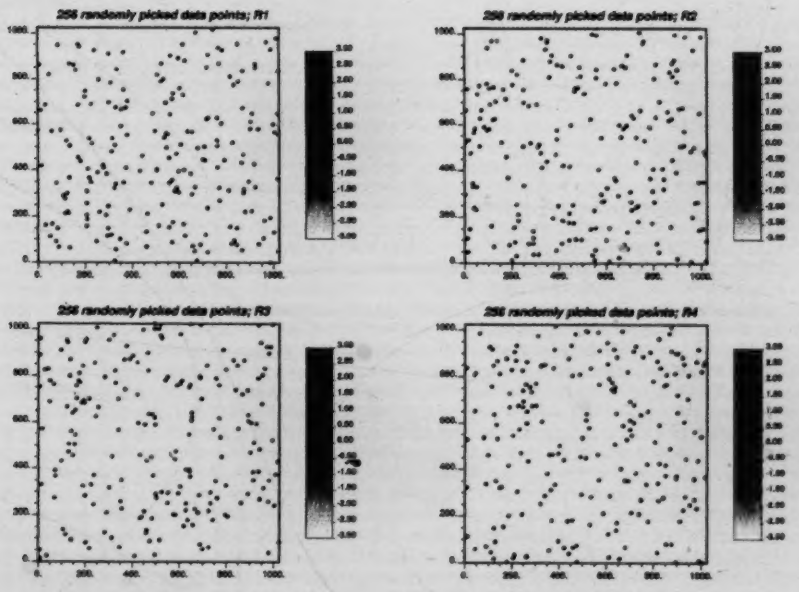


Figure 4.16. Locations of 256 of randomly picked points for different realizations (256 data locations are different) from the same reference model with field size of 1024

By using the same procedure the penalty function is calculated for various tolerance parameters. The plots of penalty function versus the lag separation distance and the tolerance ratio are shown in Figure 4.17, these plots are for just one realization of the data locations (one realization means that a fixed value for random seed number is used in drawing the data points from reference model). The conditional expectation plots for the penalty function for different number of data points and fixed field size of 1024 for different realizations (random seed numbers) are shown in Figure 4.18. The solid line shows the average over different realizations. The optimum tolerance parameters which minimize the penalty function are calculated and the associated experimental variogram and fitted variogram along with the reference variogram are plotted in Figure 4.19 for different cases of number of randomly picked data points with a fixed field size. From these plot it can be concluded that as number of data increases (with fixed field size) the number of data per unit area (or volume) will increase, therefore the optimal tolerance ratio should be decreased (Figure 4.20).

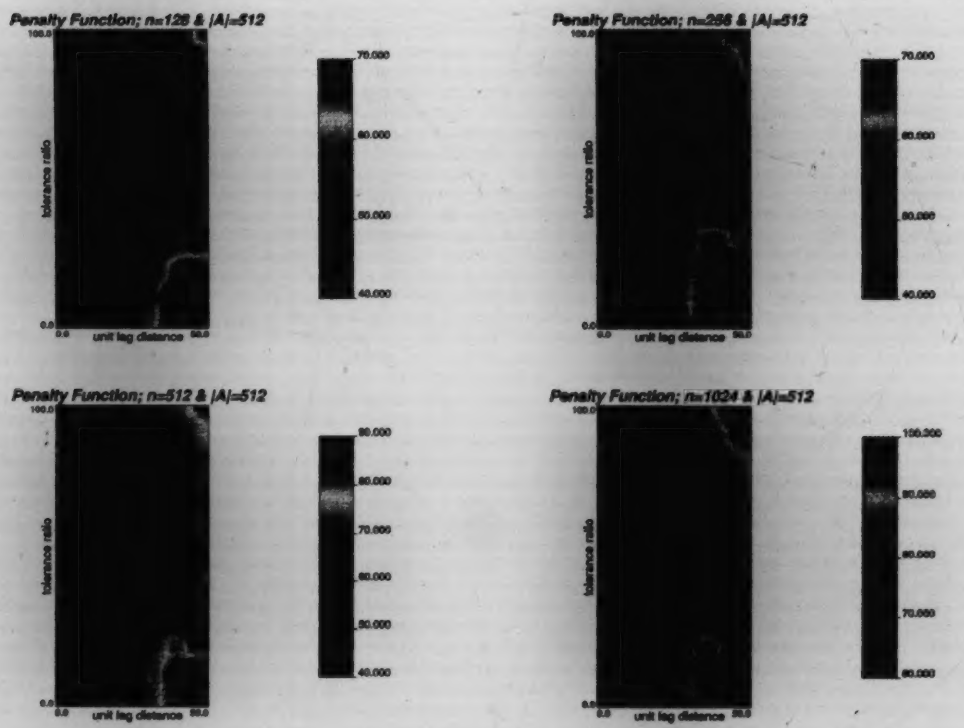

Penalty Function; n=128 & |A|=512
Penalty Function; n=256 & |A|=512
Penalty Function; n=512 & |A|=512
Penalty Function; n=1024 & |A|=512
tolerance ratio
unit lag distance

Figure 4.17. Penalty function in the case of different number of data and a fixed field size

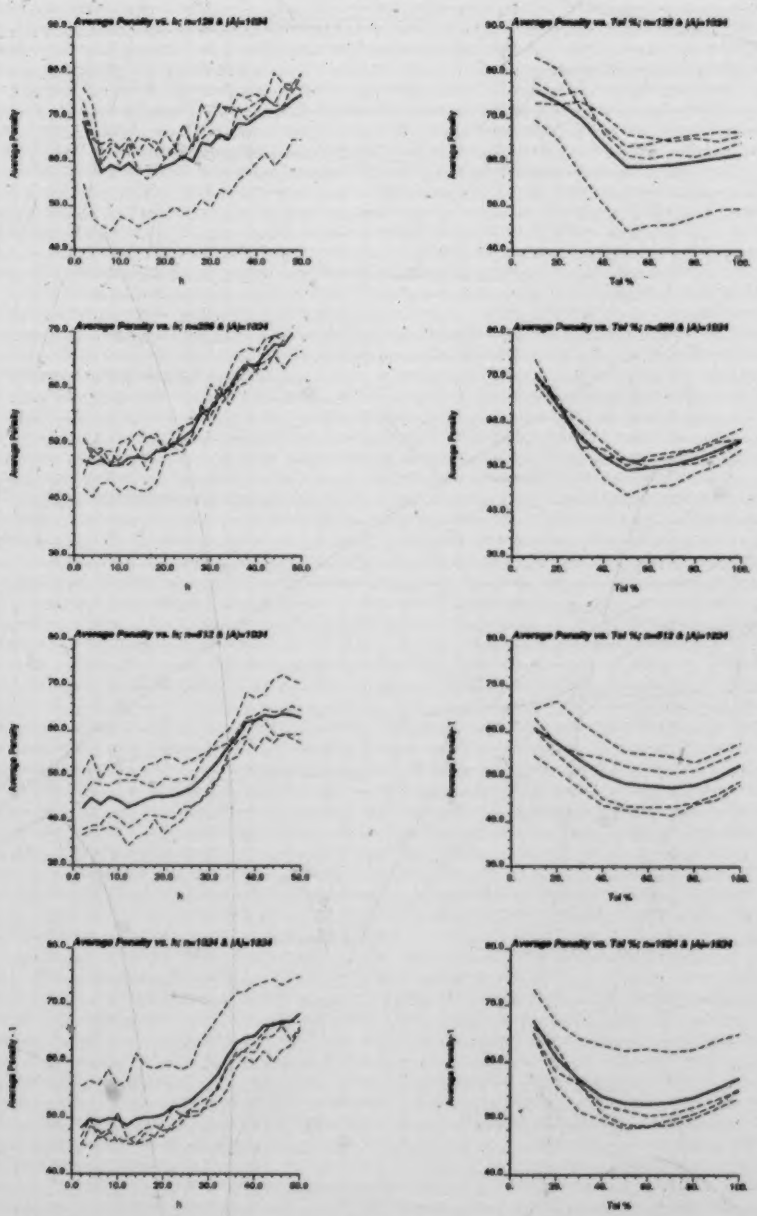


Figure 4.18. Conditional expectation of the average penalty versus the unit lag distance (left) and the tolerance ratio (right) for different number of picked points and a fixed field size of 1024, the solid line is the average over different realizations (the dashed lines)

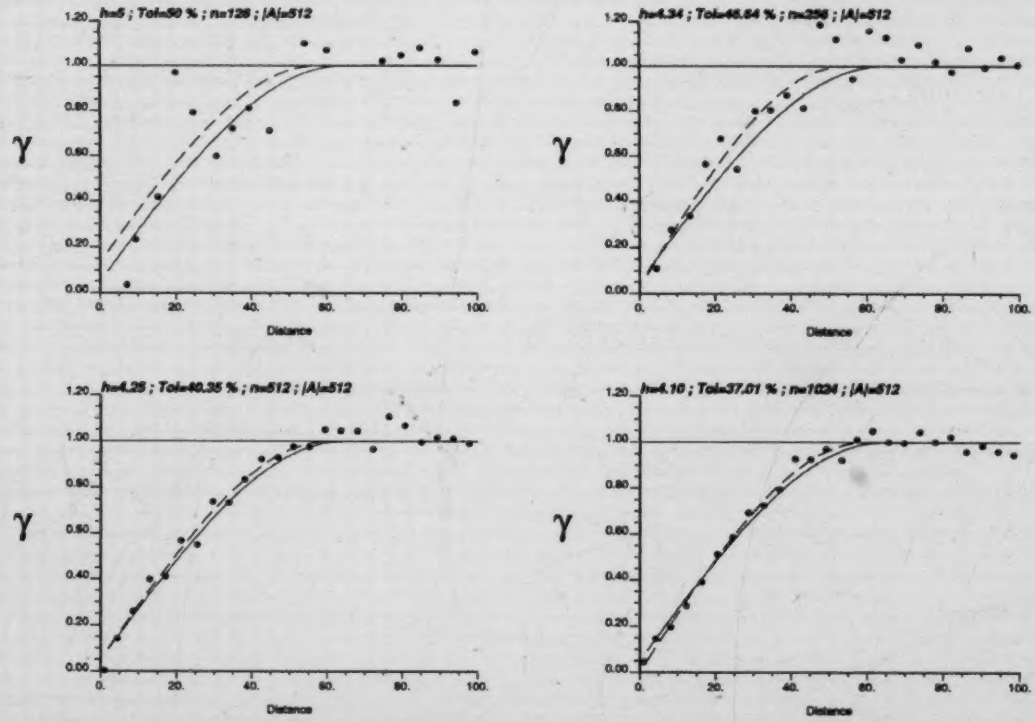


Figure 4.19. Fitted variogram (dashed line) versus the reference variogram (solid line) in the case of different number of data (n=128, 256, 512 and 1024) and a fixed field size ($|A| = 512$) using optimal tolerance parameters. Dots are the experimental variogram points.

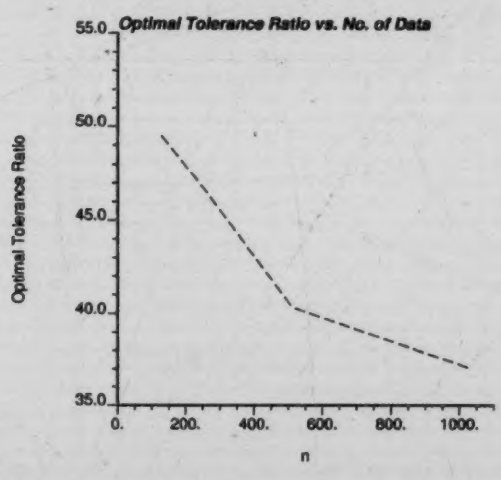


Figure 4.20. Optimal tolerance ratio versus the number of data

4.2 Real example

The methodology for improving variogram calculation will be applied for Amoco data. Amoco 2D and 3D data set are considered for implementing the methodology. The spatial variable is the porosity of the reservoir. The 2D map of the well locations along with the data values for porosity is shown in Figure 4.21. There are 62 wells in the reservoir. The 2D field size of the reservoir is 10400x10400. The histogram and cumulative distribution function of the porosity data are shown in Figure 4.22.

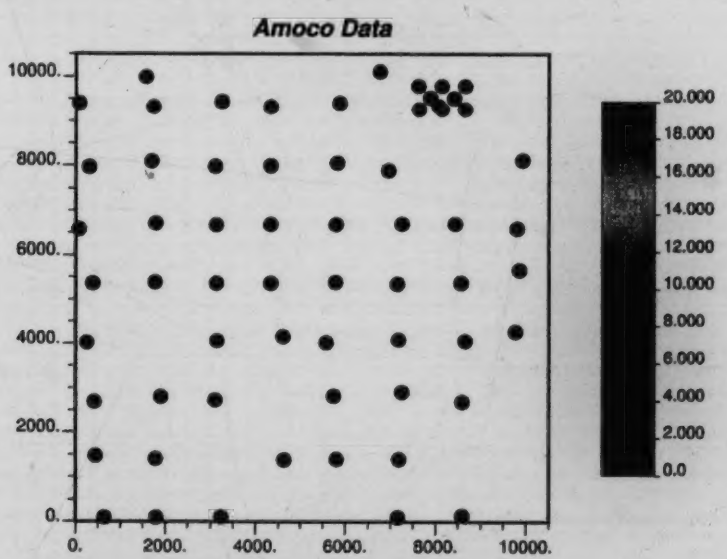


Figure 4.21. Location map of the 2D porosity data

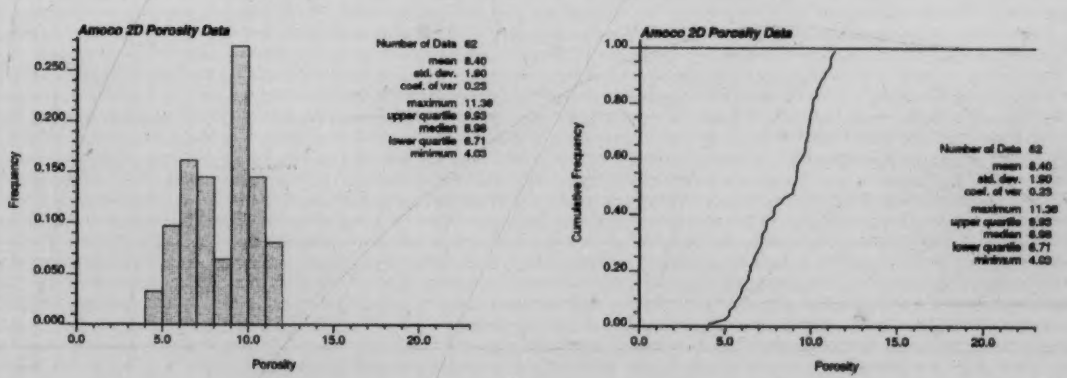


Figure 4.22. Location map of the 2D porosity data

The nominal spacing between data is 1320.80:

$$\text{nominal spacing} = \sqrt{\frac{A}{n}} = \sqrt{\frac{10400 \times 10400}{62}} = 1320.80$$

The histogram and the cumulative distribution function of the minimum distance between wells are shown in Figure 4.23.

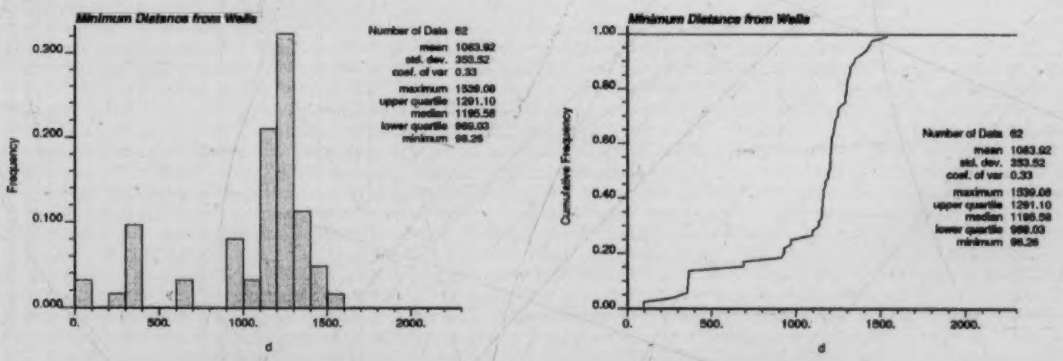


Figure 4.23. Distribution of the minimum distance between wells

The procedure that is considered to optimize the variogram tolerance parameters is as below:

1. Transforming data to normal score
2. Spatial Bootstrap; L=100, there are 100 values at each well (62 wells)
3. Calculating experimental variogram for each realization for different tolerance parameters:

$$h = 80 \times ilag \quad ; \quad ilag = 1,2,3,\ldots,25$$

$$Tol\ \% = \frac{h_{tol}}{h} \times 100 = 10 \times itol \quad ; \quad itol = 1,2,3,\ldots,10$$

$$n_{lag} = \text{int}\left(\frac{|A|}{2h}\right) = \text{int}\left(\frac{10400}{2 \times 80 \times ilag}\right) = \text{int}\left(\frac{65}{ilag}\right) \quad ; \quad ilag = 1,2,3,\ldots,25$$

$25 \times 10 \times 100 = 25,000$ total variograms should be calculated for different tolerance parameters.

4. Fitting variogram by using `varfit` program for each calculated experimental variogram (25,000 variograms).

5. Defining an assumed to be true variogram with spherical structure and nugget effect of 20 % and the range of one third of the field size as a reference variogram for calculation of the penalty function:

$$a \approx \frac{1}{3}|A| \quad \Rightarrow \quad a \approx 3500$$

$$\gamma(\mathbf{h}) = 0.2 + 0.8 \times Sph_{a=3500}(\|\mathbf{h}\|)$$

6. Calculating penalty function for each realization and each set of tolerance parameters. At last the penalty function should be averaged over 100 realizations.

   There are total 250 points (250 different set of parameters); the plot is smoothed by using inverse distance averaging method and shown in Figure 4.24. The minimum of the penalty function happens at h=1098 and Tol=10 %.

   The angle tolerance is also optimized. For optimization of angle tolerance, the optimized unit lag separation distance and the lag tolerance (h=1098, h_{tol}=109.8) are used. In this case the average penalty function is calculated for different angle tolerance (5, 10,... ,90 degrees). The minimum of the penalty function happens at a_{tol}=40 degrees. Figure 4.25 shows the plot of penalty function versus the angle tolerance.

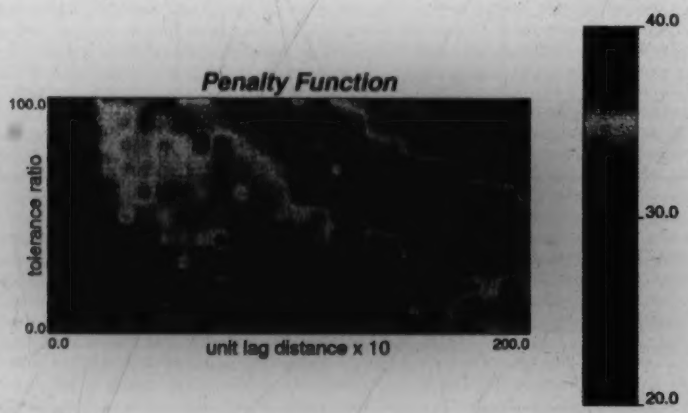


Figure 4.24. Plot of penalty function as a function of unit lag distance and the tolerance ratio

For choosing the variogram model of $\gamma(\mathbf{h}) = 0.2 + 0.8 \times Sph_{a=3500}(\|\mathbf{h}\|)$, a sensitivity analysis is performed on the selection of the nugget effect and the variogram range by using the Amoco 2D data. First the range of correlation of the assumed to be the reference variogram is fixed to the 30 percent of the field size and then the nugget effect is changed from 0 percent to 100 percent. The plots of the penalty function for this case is shown in Figure 4.26. In second case the nugget effect is fixed to 0 percent and now the range of correlation of the variogram is changed from 10 percent to 100 percent of the field size. The plot of the penalty function for this case is shown in Figure 4.27. As it can be seen from Figure 4.26 and 4.27, the low value region for the penalty function is approximately unchanged when the nugget effect varies from 0 percent to 30 percent and the range of correlation (reference variogram range) varies between 10 percent and 100 percent of the field size. In the case of fixed reference variogram range and the nugget effect between 40 percent to 100 percent, the locations of the low values are not the same as other plots. Therefore for reference variogram range, one-third of the field size is a reasonable value for reference variogram.

The Amoco 3D data set is also considered to obtain the optimized vertical bandwidth. The proportional stratigraphic coordinate is used for z-direction. The stratigraphic z-coordinate is not multiplied by the average thickness of the layer therefore it is between 0 and 1. The optimized vertical band width should be multiplied by the average thickness to have the actual value. Other tolerance parameters are assumed to be constant when the vertical bandwidth is optimized. They are assumed to be a reasonable value. The vertical angle tolerance of 15 degree, the unit lag distance was considered to be the average of the data spacing in stratigraphic coordinate (in original coordinate, since the Amoco data is well log data, the vertical data spacing is constant but when the coordinate is transformed stratigraphically, the data spacing varies). Figure 4.28 shows the plot of the penalty function versus the vertical bandwidth. It can be seen from plot that the optimized vertical bandwidth happens at 0.002 in stratigraphic unit.

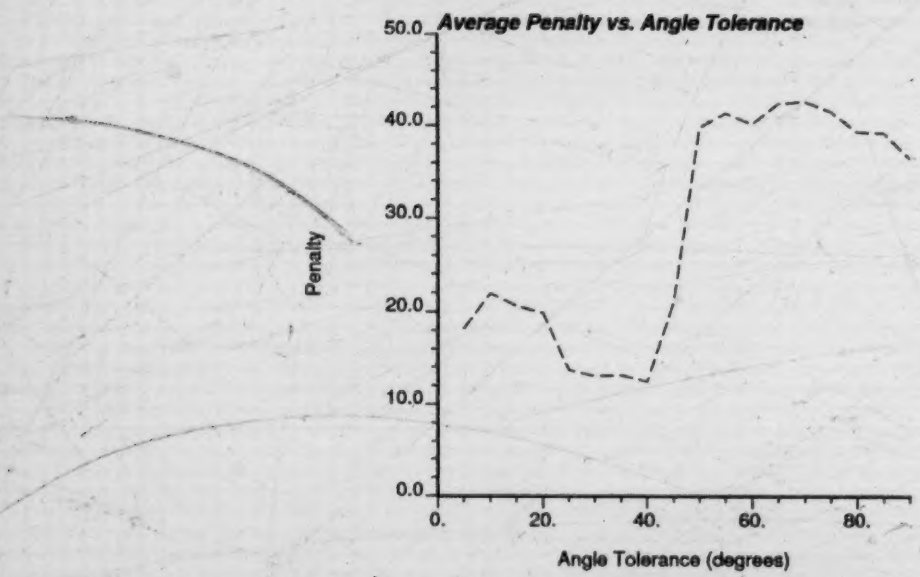


Figure 4.25. Plot of penalty function as a function of angle tolerance for Amoco 2D data set

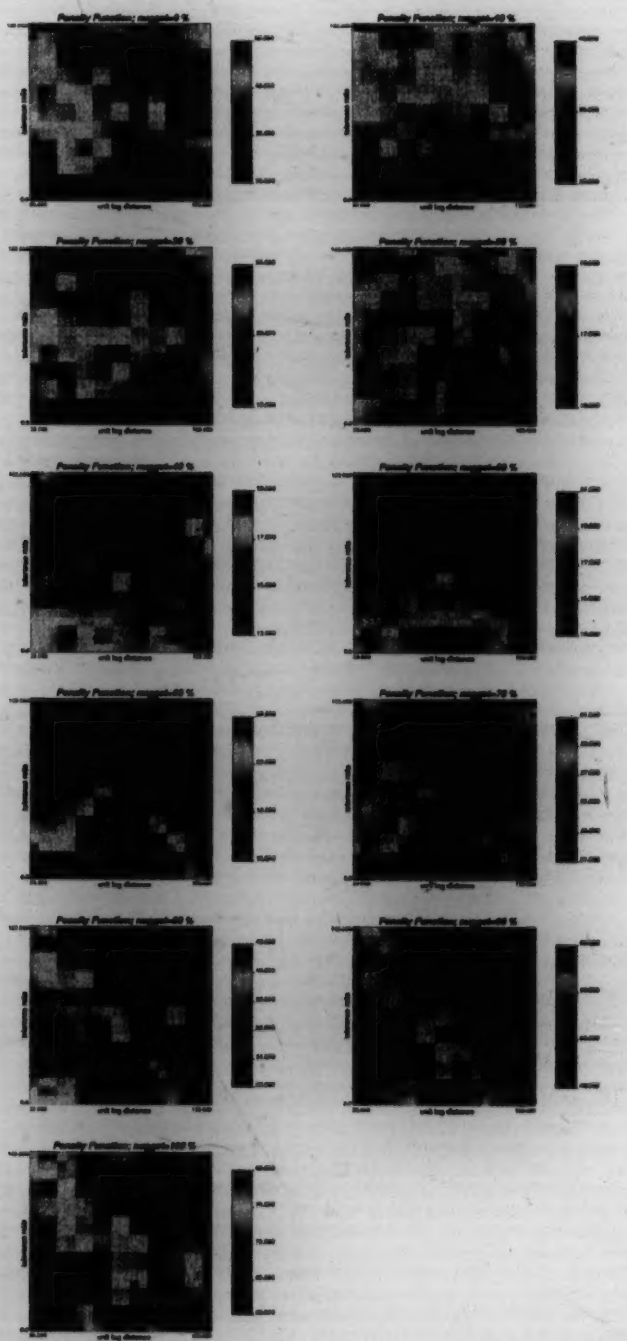


Figure 4.26. Plot of penalty function as a function of unit lag distance and the tolerance ratio for different nugget effect values for reference variogram model used in penalty function calculation for Amoco 2D data; the reference variogram range is set to be 30 percent of the field size

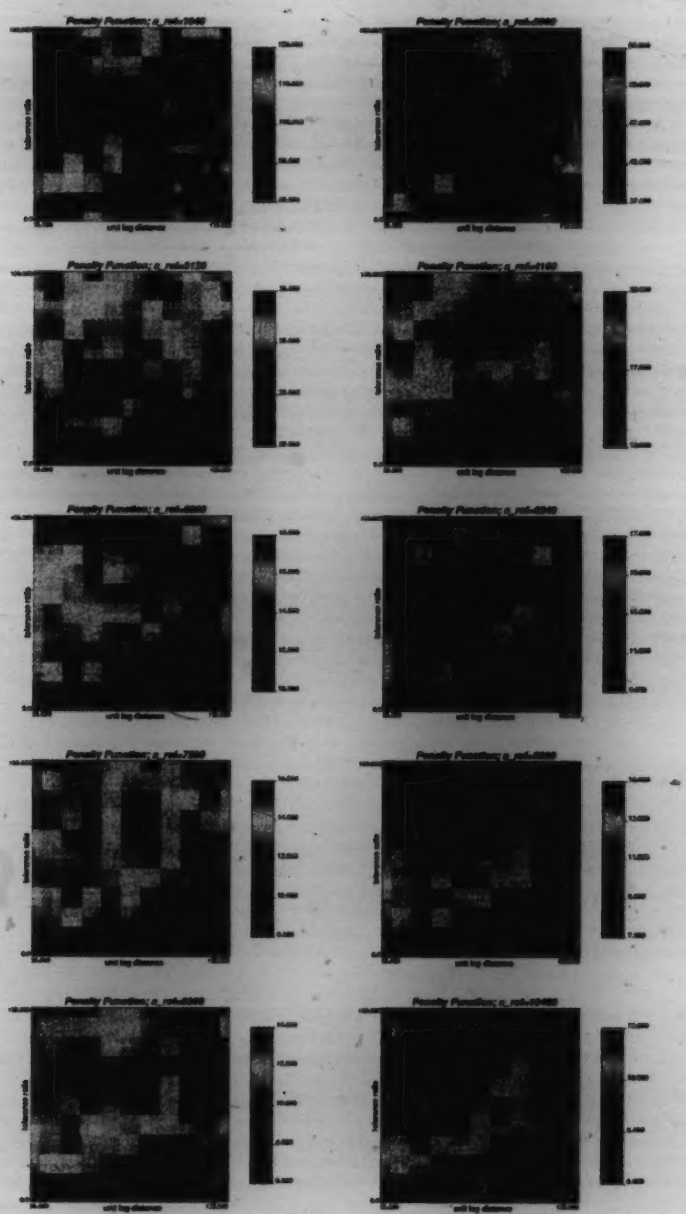


Figure 4.27. Plot of penalty function as a function of unit lag distance and the tolerance ratio for different reference variogram range for variogram model used in penalty function calculation for Amoco 2D data; the nugget effect is set to be 0 percent

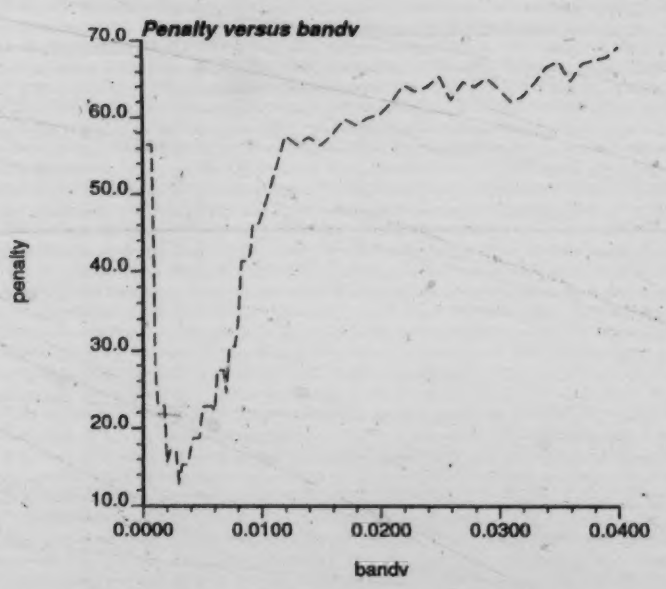


Figure 4.28. Plot of penalty function as a function of vertical bandwidth for Amoco 3D data set, the vertical bandwidth is a value between 0 and 1 because it is in proportional stratigraphic unit which is a fraction

Chapter 5

Conclusion and Future Work

5.1 Optimal tolerance parameters for variogram calculation

A methodology for optimizing tolerance parameters for variogram calculation has been proposed. The methodology is straightforward and the workflow was presented. The optimality is defined as how close the fitted variogram to the true variogram. Basically, the idea is to minimize the error between the fitted variogram and an assumed to be true variogram. The assumed to be true variogram was considered to be one structure isotropic spherical variogram with no nugget effect. A simple sensitivity on this selection was shown. The sensitive parameters that were checked are the nugget effect (ranging from 0 to 100 percent while the reference variogram range is constant) and the reference variogram range (ranging from 10 percent to 100 percent of the field size while the nugget effect is fixed). This example showed that the tolerance parameters are approximately the same in these cases for a given data set. Based on the experimental variogram point a variogram can be fitted and it can be compared to the assumed to be true variogram and a penalty can be defined based on the difference between these two variogram function. To consider better the uncertainty associated to the variogram points a spatial bootstrap is used, then the penalty value at each set of tolerance parameters is the averaged of the penalty values for different realizations. Spatial bootstrap uses unconditional simulation and assumes all parameters are fixed.

Fitting the experimental variogram reduces the uncertainty in the variogram at each lag distance. It was shown by using spatial bootstrap. After minimizing the penalty function and getting the optimized tolerance parameters the experimental variogram should be

calculated by using these tolerance parameters. Basically introducing tolerance parameters in variogram calculation overestimates the range of correlation in minor direction and underestimates the range of correlation in major direction. This amount of underestimation or overestimation was determined analytically. The correction factor in major, minor (in both 2D and 3D data set) and vertical direction (in 3D) should be determined and then the lag distances should be rescaled by using these correction factors and then refitted.

5.2 Future work

A wide variety of optimization techniques might be used for optimizing tolerance parameters. Spatial bootstrap was presented to assess uncertainty in tolerance parameters optimization. Conditional finite domain (CFD) approach (Babak and Deutsch, 2007) can be used instead of spatial bootstrap to parameter the uncertainty in the statistics of interest (i.e. penalty function). There are some limitations in using spatial bootstrap comparing to CFD approach. First, it allows only the quantification of uncertainty of order one but CFD approach allows the quantification of uncertainty of any order. Second, it does not directly include the conditioning data and the size of the domain of interest into uncertainty assessment but CFD does and thirdly it does not account for all possible data in the area of interest but CFD does.

Bibliography

Abramowitz, M. & Stegun, I. A., *Handbook of Mathematical Functions*, Dover, New York, 1965

Babak, O. and Deutsch C.V., 2007, "Reserves Uncertainty Calculation Accounting for Parameter Uncertainty", *Canadian International Petroleum Conference*, paper no. 2007-099

Deutsch, C.V. and Journel, A.G., 1998: *GSLIB - Geostatistical software library and users guide*. Oxford University press, 2nd Edition

Deutsch, C.V., 2002: *Geostatistical Reservoir Modeling*. Oxford University press

Deutsch, C.V., "A Statistical Resampling Program for Correlated Data: `Spatial_Boostrap`", *Centre for Computational Geostatistics Annual Report 6*, Department of Civil and Environmental Engineering, University of Alberta, 2004, 9 pp.

Goovaerts, P., *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, 1997, 483pp.

Gringarten, E. and Deutsch, C.V., "Teacher's Aid Variogram Interpretation and Modeling", *Mathematical Geology*, 33 (4), 2001, p. 507-534.

Journel, A.G. and Hujibregts, C.J., *Mining Geostatistics*, Academic Press Limited, 1978, 600 pp.

Journel, A.G., *Fundamentals of Geostatistics in Five Lessons*, Volume 8 Short Course in Geology, American Geophysical Union, Washington, DC, 1989, 40 pp.

Kelkar M., Perez G., 2002: *Applied Geostatistics for Reservoir Characterization*. Society of Petroleum Engineers

Neufeld, C. T. and Deutsch, C. V., "Developments in Semiautomatic Variogram Fitting", *Centre for Computational Geostatistics Annual Report 6*, Department of Civil and Environmental Engineering, University of Alberta, 2004, 12 pp.

Pardo-Igúzquiza, E. and Dowd, P., "Variance-Covariance Matrix of the Experimental Variogram: Assessing Variogram Uncertainty", *Mathematical Geology*, 33 (4), 2001, p. 397-419.

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., Numerical Recipes in Fortran 77: the Art of Scientific Computing, Second Edition, 1992.

Appendix A

In this appendix the proof for 3D tolerance ratio which was explained in chapter 3 will be presented.

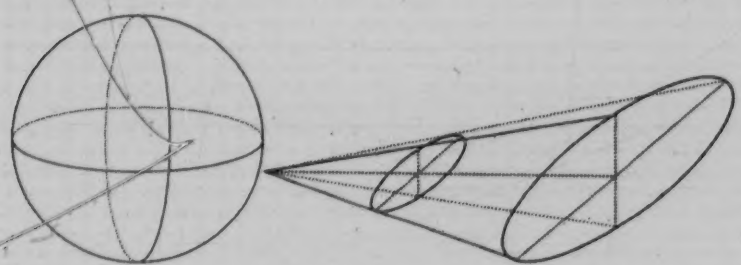
Tolerance ratio in 3D

Case I: $h \le h^{*}_{ver} - h_{tol}$

$$h^{*}_{ver} = \frac{b_{ver}}{\sin a^{v}_{tol}}$$

$V_3\left(h, a^{h}_{tol}, a^{v}_{tol}\right)$ is the half of the volume of the region which is inside both the sphere with a radius of h and an elliptic cone with angles of a^{h}_{tol}, a^{v}_{tol}. Analytically it is,

$$D: \begin{cases} x^2 + y^2 + z^2 \le h^2 \\ x^2 \cot^2 a^{h}_{tol} + y^2 \cot^2 a^{v}_{tol} \le z^2 \\ z \ge 0 \end{cases} \quad \Rightarrow \quad V_3\left(h, a^{h}_{tol}, a^{v}_{tol}\right) = \iiint_D dxdydz$$



To solve the above triple integral the coordinates should be changed to spherical coordinates, therefore:

$$D:\begin{cases}\rho \le h \\ \cos^2\theta\cot^2 a_{tol}^h + \sin^2\theta\cot^2 a_{tol}^v \le \cot^2\phi\end{cases} \Rightarrow V_3\left(h, a_{tol}^h, a_{tol}^v\right) = \iiint_D \rho^2 \sin\phi d\rho d\phi d\theta$$

And finally the triple integral would be

$$V_3\left(h, a_{tol}^h, a_{tol}^v\right) = \int_0^{2\pi}\int_0^{\phi_1}\int_0^{h} \rho^2 \sin\phi d\rho d\phi d\theta = \frac{h^3}{3}\int_0^{2\pi}\left(1-\cos\phi_1\right)d\theta$$

Where ϕ_1 should be satisfied in below equation:

$$\cot^2\phi_1 = \cos^2\theta\cot^2 a_{tol}^h + \sin^2\theta\cot^2 a_{tol}^v \Rightarrow \cos\phi_1 = \left(\frac{\cos^2\theta\cot^2 a_{tol}^h + \sin^2\theta\cot^2 a_{tol}^v}{1+\cos^2\theta\cot^2 a_{tol}^h + \sin^2\theta\cot^2 a_{tol}^v}\right)^{\frac{1}{2}}$$

Therefore the formula for the volume becomes:

$$V_3\left(h, a_{tol}^h, a_{tol}^v\right) = \frac{2\pi}{3}\left[1 - I_1\left(a_{tol}^h, a_{tol}^v\right)\right]h^3$$

Where:

$$I_1\left(a_{tol}^h, a_{tol}^v\right) = \frac{1}{2\pi}\int_0^{2\pi}\cos\phi_1 d\theta = \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{\cos^2\theta\cot^2 a_{tol}^h + \sin^2\theta\cot^2 a_{tol}^v}{1+\cos^2\theta\cot^2 a_{tol}^h + \sin^2\theta\cot^2 a_{tol}^v}\right)^{\frac{1}{2}} d\theta$$

The tolerance ratio is defined as the

$$Tol\,\% = \frac{V_3\left(h+h_{tol}, a_{tol}^h, a_{tol}^v\right) - V_3\left(h-h_{tol}, a_{tol}^h, a_{tol}^v\right)}{V(2h)}$$

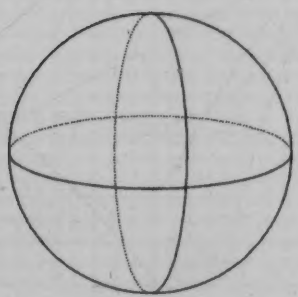
$$= \frac{\frac{2\pi}{3}\left[1 - I_1\left(a_{tol}^h, a_{tol}^v\right)\right]\left[\left(h+h_{tol}\right)^3 - \left(h-h_{tol}\right)^3\right]}{\frac{2\pi}{3}(2h)^3}$$

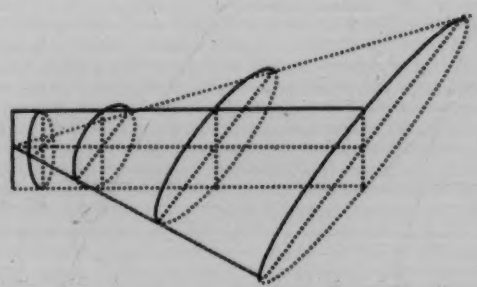
$$= \frac{\left[1 - I_1\left(a_{tol}^h, a_{tol}^v\right)\right]\left[3h^2 + h_{tol}^2\right]h_{tol}}{4h^3} = \frac{1}{4}\left[1 - I_1\left(a_{tol}^h, a_{tol}^v\right)\right]\left(\frac{h_{tol}}{h}\right)\left[3 + \left(\frac{h_{tol}}{h}\right)^2\right] \Rightarrow$$

Case II: $h_{ver}^{\bullet} + h_{tol} < h \le h_{hor}^{\bullet} - h_{tol}$

$V_4\left(h, b_{ver}, a_{tol}^h\right)$ is the volume of the region within the sphere with a radius of a h and the shape with a cross section of an ellipse, one of the main axis of this ellipse has a constant length of b_{ver}. Analytically it is,

$$D:\begin{cases} x^2+y^2+z^2 \le h^2 \\ x^2\cot^2 a_{tol}^h + \dfrac{y^2 z^2}{b_{ver}^2} \le z^2 \\ z \ge 0 \end{cases} \quad \Rightarrow \quad V_4\left(h, b_{ver}, a_{tol}^h\right) = \iiint_D dxdydz$$





$$V_4\left(h, b_{ver}, a_{tol}^h\right) = V_1^* + V_2^*$$

$$V_1^* = \int_0^{h\cos a_{tol}^h} \pi b_{ver} z \tan a_{tol}^h dz = \pi b_{ver} \tan a_{tol}^h \frac{h^2\cos^2 a_{tol}^h}{2} = \frac{\pi \sin\left(2a_{tol}^h\right)}{4} b_{ver} h^2$$

$$V_2^* = \frac{1}{2}\left[V^* - V_{\text{elliptic cylinder}}\left(b_{ver}, h\sin a_{tol}^h, 2h\cos a_{tol}^h\right)\right] = \frac{1}{2}\left[V^* - 2\pi b_{ver}\sin a_{tol}^h \cos a_{tol}^h h^2\right]$$

$$= \frac{V^*\left(h, b_{ver}, a_{tol}^h\right)}{2} - \frac{\pi\sin\left(2a_{tol}^h\right)}{2} b_{ver} h^2$$

$$\Rightarrow \quad V_4\left(h, b_{ver}, a_{tol}^h\right) = \frac{V^*\left(h, b_{ver}, a_{tol}^h\right)}{2} - \frac{\pi\sin\left(2a_{tol}^h\right)}{4} b_{ver} h^2$$

Where $V^*\left(h, b_{ver}, a_{tol}^h\right)$ is the volume within the region inside both of the elliptic cylinder (with minor and major axis of $b_{ver}, h\sin a_{tol}^h$) and the sphere with a radius of h.

Analytically it is:

$$D:\begin{cases} x^2+y^2+z^2 \le h^2 \\ \dfrac{x^2}{h^2\sin^2 a_{tol}^h}+\dfrac{y^2}{b_{ver}^2} \le 1 \end{cases} \quad \Rightarrow \quad V^*\left(h,b_{ver},a_{tol}^h\right)=\iiint_D dxdydz$$

In cylindrical coordinates it reduces to

$$D:\begin{cases} r^2+z^2 \le h^2 \\ r^2 \le \dfrac{1}{\dfrac{\cos^2\theta}{h^2\sin^2 a_{tol}^h}+\dfrac{\sin^2\theta}{b_{ver}^2}} \end{cases} \quad \Rightarrow \quad V^*\left(h,b_{ver},a_{tol}^h\right)=\iiint_D rdzdrd\theta \quad \Rightarrow$$

$$V^*\left(h,b_{ver},a_{tol}^h\right)=\int_0^{2\pi}\int_0^{r_1}\int_{-z_1}^{z_1} rdzdrd\theta$$

Where $z_1=\sqrt{h^2-r^2}$ and $r_1=\left(\dfrac{\cos^2\theta}{h^2\sin^2 a_{tol}^h}+\dfrac{\sin^2\theta}{b_{ver}^2}\right)^{\frac{-1}{2}}$, therefore;

$$V^*\left(h,b_{ver},a_{tol}^h\right)=\int_0^{2\pi}\int_0^{r_1}\int_{-z_1}^{z_1} rdzdrd\theta=2\int_0^{2\pi}\int_0^{r_1} rz_1 drd\theta=2\int_0^{2\pi}\int_0^{r_1} r\sqrt{h^2-r^2}\,drd\theta$$

$$=2\int_0^{2\pi}\left[\frac{-1}{3}\left(h^2-r^2\right)^{\frac{3}{2}}\right]_0^{r_1} d\theta=\frac{2}{3}\int_0^{2\pi}\left[h^3-\left(h^2-r_1^2\right)^{\frac{3}{2}}\right]d\theta$$

$$\Rightarrow \quad V^*\left(h,b_{ver},a_{tol}^h\right)=\frac{4\pi}{3}\left[1-I_2\left(h,b_{ver},a_{tol}^h\right)\right]h^3$$

where

$$I_2\left(h,b_{ver},a_{tol}^h\right)=\frac{1}{2\pi}\int_0^{2\pi}\left[1-\left(\frac{r_1}{h}\right)^2\right]^{\frac{3}{2}}d\theta=\frac{1}{2\pi}\int_0^{2\pi}\left[\frac{\csc^2 a_{tol}^h\cos^2\theta+\dfrac{h^2}{b_{ver}^2}\sin^2\theta-1}{\csc^2 a_{tol}^h\cos^2\theta+\dfrac{h^2}{b_{ver}^2}\sin^2\theta}\right]^{\frac{3}{2}}d\theta$$

By knowing $V^*\left(h,b_{ver},a_{tol}^h\right)$, the formula for $V_4\left(h,b_{ver},a_{tol}^h\right)$ reduces to

$$V_4\left(h,b_{ver},a_{tol}^h\right)=\frac{V^*\left(h,b_{ver},a_{tol}^h\right)}{2}-\frac{\pi\sin\left(2a_{tol}^h\right)}{4}b_{ver}h^2$$

$$=\frac{2\pi}{3}\left[1-I_2\left(h,b_{ver},a_{tol}^h\right)\right]h^3-\frac{\pi\sin\left(2a_{tol}^h\right)}{4}b_{ver}h^2$$

The tolerance ratio is defined as the

$$Tol\,\% = \frac{V_4\left(h + h_{tol}, b_{ver}, a_{tol}^{h}\right) - V_4\left(h - h_{tol}, b_{ver}, a_{tol}^{h}\right)}{V(2h)}$$

Case III: $h_{ver}^{*} - h_{tol} < h \le h_{ver}^{*} + h_{tol}$

$$Tol\,\% = \frac{V_3\left(h_{ver}^{*}, a_{tol}^{h}, a_{tol}^{v}\right) - V_3\left(h - h_{tol}, a_{tol}^{h}, a_{tol}^{v}\right) + V_4\left(h + h_{tol}, b_{ver}, a_{tol}^{h}\right) - V_4\left(h_{ver}^{*}, b_{ver}, a_{tol}^{h}\right)}{V(2h)}$$

Case IV: $h > h_{hor}^{*} + h_{tol}$

The interested volume in this case can be derived by using the formula for $V_3\left(h, b_{hor}, b_{ver}\right)$, but instead of using the $\cot a_{tol}^{h}$ and $\cot a_{tol}^{v}$, we can use the equivalent of them in terms of the bandwidth which are equal to $\frac{h}{b_{hor}}$ and $\frac{h}{b_{ver}}$ respectively.

$$V_3\left(h, b_{hor}, b_{ver}\right) = \frac{2\pi}{3}\left[1 - I_3\left(h, b_{hor}, b_{ver}\right)\right]h^3$$

Where

$$I_3\left(h, b_{hor}, b_{ver}\right) = \frac{1}{2\pi}\int_0^{2\pi}\left[\frac{\frac{h^2}{b_{hor}^2}\cos^2\theta + \frac{h^2}{b_{ver}^2}\sin^2\theta - 1}{\frac{h^2}{b_{hor}^2}\cos^2\theta + \frac{h^2}{b_{ver}^2}\sin^2\theta}\right]^{\frac{3}{2}} d\theta$$

The tolerance ratio is defined as the

$$Tol\,\% = \frac{V_5\left(h + h_{tol}, b_{hor}, b_{ver}\right) - V_5\left(h - h_{tol}, b_{hor}, b_{ver}\right)}{V(2h)}$$

Case V: $h_{hor}^{*} - h_{tol} < h \le h_{hor}^{*} + h_{tol}$

$$Tol\,\% = \frac{V_4\left(h_{hor}^{*}, b_{ver}, a_{tol}^{h}\right) - V_4\left(h - h_{tol}, b_{ver}, a_{tol}^{h}\right) + V_5\left(h + h_{tol}, b_{hor}, b_{ver}\right) - V_5\left(h_{hor}^{*}, b_{hor}, b_{ver}\right)}{V(2h)}$$